



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Académico Profesional de Estadística

**Estudio y aplicación de los modelos ARFIMA en el
análisis de la serie de tiempo "Importación de bienes
de consumo en el Perú"**

TESINA

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

Genaro Humberto ARAUJO SALAS

ASESOR

Erwin KRAENAU ESPINAL

Lima, Perú

2009



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Araujo, G. (2009). *Estudio y aplicación de los modelos ARFIMA en el análisis de la serie de tiempo "Importación de bienes de consumo en el Perú"*. Tesina para optar el título profesional de Licenciado en Estadística. Escuela Académico Profesional de Estadística, Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

Este trabajo lo dedico en especial a mis queridos padres Hernán y Noemí por su apoyo incondicional, a mi familia en general y las personas que me apoyaron en todo momento.

Agradecimientos especiales
a los profesores Erwin
Kraenau Espinal e Ysela
Agüero Palacios por sus
importantes asesorías para
el desarrollo de este trabajo.

RESUMEN

Estudio y aplicación de los modelos ARFIMA en el análisis de la serie de tiempo
“Importación de bienes de consumo en el Perú”

GENARO ARAUJO SALAS

DICIEMBRE – 2009

Asesor : Erwin Kraenau Espinal
Título obtenido : Licenciado en Estadística

En el presente trabajo se presenta una aplicación de los modelos ARFIMA a la serie de tiempo “Importación de bienes de consumo en el Perú (en millones de dólares)”. En los conceptos básicos se repasa la teoría de los modelos bajo el enfoque del dominio del tiempo (modelos ARIMA) y bajo el enfoque del dominio de frecuencia, para contextualizarnos antes de entrar a la teoría de los modelos ARFIMA; se presentan herramientas alternativas para detectar la característica de memoria larga para poder decidir cuál de ellas utilizar, se desarrolla la teoría correspondiente a los modelos ARFIMA en la cual se verán los métodos para estimar el parámetro de diferenciación fraccional y luego se desarrolla la aplicación para luego obtener pronósticos, comentar los resultados y dar las conclusiones del caso.

Palabras Claves: ARFIMA
Dominio del Tiempo
Dominio de Frecuencia
Memoria Larga
Diferenciación Fraccional

ABSTRACT

Study and application of ARFIMA models in the analysis of time series "Imports of consumer goods in Peru (million dollars)."

GENARO ARAUJO SALAS

DICIEMBRE – 2009

Asesor : Erwin Kraenau Espinal
Título obtenido : Licenciado en Estadística

This report presents an application of ARFIMA models for time series "Imports of consumer goods in Peru (million dollars)." The basic concepts are reviewed the theory of models under the time-domain approach (ARIMA models) and under the frequency domain approach in order to contextualize before entering the ARFIMA model theory, also presents alternative tools to detect long memory characteristics in order to decide which one to use, the theory is developed for ARFIMA models in which will be the methods for estimating the fractional differencing parameter and later develops the application to then obtain forecasts, discuss the results and make appropriate conclusions.

Keywords: ARFIMA
Time Domain
Frequency Domain
Long Memory
Fractional Differencing

INDICE

ÍNDICE.....	viii
INTRODUCCIÓN.....	x
CAPITULO I GENERALIDADES	1
I.1 ANTECEDENTES:	1
I.2 FORMULACION DEL PROBLEMA:	2
I.3 JUSTIFICACIÓN:	2
I.4 OBJETIVOS:.....	2
I.5 IMPORTANCIA O IMPACTO:	3
I.6 VIABILIDAD O FACTIBILIDAD:	3
CAPITULO II MARCO TEÓRICO	4
II.1 CONCEPTOS BÁSICOS.....	4
II.1.1 Procesos Estocásticos.....	4
II.1.2 Procesos Estacionarios	4
II.1.3 Proceso Gaussiano	5
II.1.4 Ruido Blanco	5
II.1.5 Procesos de memoria Corta.....	5
II.1.6 Procesos de memoria larga.....	6
II.2 Análisis bajo el Enfoque del Dominio del Tiempo (Modelos AR (p), MA (q), ARMA (p,q) Y ARIMA(p,d,q))	6
II.2.1 Modelos Autoregresivos AR(p)	6
II.2.2 Modelos de Medias Móviles MA(q)	8
II.2.3 Modelos ARMA (p,q).....	10
II.3 Integración y Diferenciación	11
II.4 Modelos ARIMA (p,d,q).....	12
II.5 Análisis bajo el enfoque de dominio de frecuencia	13
II.5.1 Función Periódica	13
II.5.2 Cosenoidales y Sinusoidales.....	13
II.5.3 Armónica	14
II.5.4 Análisis de Fourier y análisis armónico	15
II.5.5 Periodograma de Schuster	16
II.5.6 Análisis espectral de un proceso estacionario	16
II.5.7 Función espectral de un modelo ARMA:.....	17

II.6 Modelos ARFIMA.....	18
II.6.1 Definición y propiedades	18
II.6.2 La función gamma y los procesos de memoria larga	25
II.6.3 Identificación de modelos ARFIMA	26
II.6.4 Dósimas para contrastar la hipótesis de memoria larga	28
II.7 Métodos de estimación para el parámetro de diferenciación fraccional “d”	31
II.7.1 Método de regresión usando la función de periodograma	31
II.7.2 Método de regresión usando la función de periodograma suavizado	34
II.7.3 Método de Máxima Verosimilitud	36
II.8 Procedimiento general para estimar los parámetros del modelo ARFIMA (p,d,q)	37
CAPITULO III APLICACIÓN DE LA SERIE “IMPORTACION DE BIENES DE CONSUMO EN EL PERU (\$ MILLONES) “(1985-2008)	38
III.1 Análisis descriptivo de la serie “Importación de Bienes de Consumo en el Perú (\$ mill)” ..	40
III.2 Identificación de memoria larga en la serie “Importación de Bienes de Consumo en el Perú (\$ mill)”	44
III.3 Análisis de la serie “Importación de Bienes de Consumo (\$ mill)” bajo los modelos ARFIMA	46
III.4 Comparación con el modelo ARIMA	50
CONCLUSIONES.....	52
BIBLIOGRAFÍA.....	53
ANEXOS	55

INTRODUCCION

Cuando se estudia series de tiempo bajo el enfoque del dominio del tiempo uno puede encontrarse con series que presentan memoria corta o memoria larga , esta característica depende de cómo se comportan las autocorrelaciones a través del tiempo; en particular las series de tiempo que presentan memoria larga requieren de un análisis cuidadoso donde no se deba perder este efecto, ya que en esta característica radica la valiosa información que una serie de este tipo pueda tener , los modelos ARIMA con los cuales se estudian las series de tiempo bajo el enfoque del dominio del tiempo tienen ciertas restricciones que hacen que una serie pierda el efecto de memoria larga al diferenciar la serie.

Las series económicas se caracterizan (la mayoría) por tener este efecto pues sus datos guardan una relación importante entre valores distantes. Otra característica que se debe tener en cuenta, es como la serie responde en el futuro ante posibles shocks, pues bajo el análisis de los modelos ARIMA la serie es persistente y no hay reversión a la media.

En el presente trabajo se mostrará porque es conveniente modelar la importación de bienes de consumo en el Perú usando los modelos ARFIMA y no los tradicionales modelos ARIMA.

Hoy en día los modelos ARFIMA son muy usados no solamente para modelar únicamente la media de una serie, sino también en conjunto con otros modelos que modelan la varianza de la serie (como los ya conocidos modelos FIGARCH).

CAPITULO I

GENERALIDADES

I.1 ANTECEDENTES:

El desarrollo de los modelos ARFIMA empezó con los trabajos de Hurst (1951) en el campo de la Hidrología, sin embargo los trabajos aplicados a otras ciencias en particular económicas se dan a partir de los trabajos de Granger (1980) y Granger y Joyeux (1980); Ellos sugieren que al diferenciar una serie hasta volverla estacionaria puede tener consecuencias negativas, en particular en muchas series económicas que en apariencia son no estacionarias con espectro no acotado en el origen. Lo tradicional es diferenciar para conseguir varianza finita, sin embargo, la serie diferenciada se convierte a menudo en una serie con espectro nulo en el cero eliminando de la serie original la componente de las bajas frecuencias tan importantes en las predicciones a largo plazo, por ello Granger (1980), Granger y Joyeux (1980) y Hosking (1981) proponen una clase de procesos intermedios en el que el orden de integración sea fraccionario, se trata de los procesos autoregresivos y de medias móviles fraccionalmente integrados ARFIMA (p, d, q) donde “ d ” es un número real al admitir que el orden de integración “ d ” sea un número no entero, estos modelos actúan como un nexo entre los procesos con raíces unitarias ARIMA ($d=1$) y los procesos ARMA estacionarios que no requieren integración ($d=0$).

En la actualidad existe una serie de trabajos aplicativos referidos a los modelos ARFIMA, sobre todo en el campo de la economía donde las series de tiempo tienen como principal característica el efecto de memoria larga y donde los componentes a largo plazo tienen un papel muy importante en el análisis ya que en estas componentes radica la información más importante para el análisis.

I.2 FORMULACION DEL PROBLEMA:

La aplicación de modelos ARFIMA se da en muchas áreas como la hidrología, economía y otros campos, en particular en el área de la economía las series se caracterizan por su volatilidad y efecto de memoria larga; cabe hacernos la siguiente pregunta: ¿Por qué utilizar los modelos ARFIMA y no los tradicionales modelos ARIMA para analizar la serie de tiempo económica “Importación de bienes de consumo en el Perú”?

I.3 JUSTIFICACIÓN:

Revisando estudios pasados hemos podido apreciar la ventaja que se tiene al analizar una serie de tiempo que presenta memoria larga cuando utilizamos modelos ARFIMA a comparación de los modelos tradicionales ARIMA, sin embargo el hecho de hacer otra comparación surge del deseo de mostrar algunas herramientas adicionales que aún no se han mostrado en otros estudios al analizar series económicas con modelos ARFIMA; así mismo el deseo del autor de presentar los modelos ARFIMA como alternativa en el estudio de series temporales en particular económicas con resultados que quedaran como material de consulta para estudiantes de pregrado y público en general interesados en el tema.

I.4 OBJETIVOS:

GENERALES

Estudiar los modelos ARFIMA y aplicarlo en el análisis de una serie de tiempo con memoria larga.

ESPECIFICOS

- Estudiar los modelos ARIMA y ARFIMA
- Realizar el análisis de la serie de tiempo “importación de bienes de consumo” mediante los modelos ARFIMA para obtener pronósticos.
- Presentar pruebas alternativas para el parámetro de diferenciación fraccional.

I.5 IMPORTANCIA O IMPACTO:

Este trabajo contribuirá al área de la ciencia estadística, en particular al análisis de series temporales económicas bajo el enfoque del dominio del tiempo. Servirá de material de consulta para posteriores investigaciones de series económicas.

I.6 VIABILIDAD O FACTIBILIDAD:

La literatura disponible sobre modelos ARFIMA ha ido incrementándose a partir de fines del siglo pasado a partir de los trabajos de Granger (1980) y Granger y Joyeux (1980), diversas aplicaciones sobre todo en el campo de la economía, teniendo en cuenta que en la actualidad la cantidad de material bibliográfico acerca del tema es mayor (la mayoría en papers y en ingles), así como los programas estadísticos de los cuales se tiene fácil acceso como el programa R por ser un software de libre uso.

CAPITULO II

MARCO TEÓRICO

II.1 CONCEPTOS BÁSICOS

Antes de entrar a la teoría de los modelos ARFIMA, será necesario revisar algunos conceptos básicos, que nos servirán a lo largo del estudio, algunos conceptos se dan a continuación:

II.1.1 Procesos Estocásticos

Se define un proceso estocástico como una familia de variables aleatorias observadas a través del espacio, el tiempo, etc, cada una de las variables aleatorias tiene una función de distribución poblacional definida y pueden o no estar correlacionadas entre si.

En particular en series de tiempo estas variables aleatorias. son observadas a través del tiempo t $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ $t \in \langle -\infty, \infty \rangle$. Un proceso estocástico puede estar caracterizado de dos maneras mediante su función de distribución poblacional o mediante sus momentos.

II.1.2 Procesos Estacionarios

Se dice que un proceso es estacionario en sentido estricto, si al hacer un mismo desplazamiento "r" a todas las variables aleatorias. resulta que la función de distribución poblacional no varía, es decir:

$$f(y_{t_1}, y_{t_2}, \dots, y_{t_n}) = f(y_{t_1+r}, y_{t_2+r}, \dots, y_{t_n+r})$$

Se dice que un proceso es estacionario en sentido amplio o de primer orden si la media μ del proceso y_t permanece constante $\forall t \in \langle -\infty, \infty \rangle$.

$$E(y_t) = \mu$$

Un proceso es estacionario en sentido amplio o de segundo orden si la media μ y la varianza σ^2 (finita) permanecen constantes a través del tiempo $\forall t \in \langle -\infty, \infty \rangle$.

$$E(y_t) = \mu$$

$$\text{Var}(y_t) = \sigma^2$$

Todo proceso estacionario en sentido estricto es estacionario en sentido amplio, pero no se cumple el recíproco.

II.1.3 Proceso Gaussiano

Un proceso Gaussiano se define como una sucesión de variables aleatorias independientes e idénticamente distribuidas cada una de ellas con distribución de probabilidad normal.

II.1.4 Ruido Blanco

Una colección de variables aleatorias incorrelacionadas entre si se le denomina ruido blanco, el término blanco viene del sentido “luz blanca” que indican todas las posibles oscilaciones periódicas que están presentes con igual fuerza.

En particular una colección de variables aleatorias con media cero y varianza σ^2 finita se le denomina ruido blanco gaussiano.

II.1.5 Procesos de memoria Corta

Son aquellos que se caracterizan cuando sus autocovarianzas (o dependencias) entre y_{t_1} y y_{t+k_1} decrecen rápidamente cuando k tiende al infinito, es decir que

cuando el número de retardos aumenta la función de autocorrelación decrece exponencialmente y es limitada, además sus autocorrelaciones son sumables y su función de densidad espectral se elimina en las frecuencias bajas un ejemplo en particular es el modelo ARMA (1,0).

II.1.6 Procesos de memoria larga

Estos procesos se caracterizan a diferencia de los de memoria corta cuando sus autocovarianzas entre y_{t_1} y y_{t+k_1} decrecen lentamente cuando k tiende al infinito. Su función de densidad espectral no está acotada para frecuencias bajas y sus autocorrelaciones no son sumables es decir:

$$\sum |\rho_k| = \infty$$

El término “proceso” y “modelo” son equivalentes siendo el término más formal y utilizado el de “modelo”

II.2 Análisis bajo el Enfoque del Dominio del Tiempo (Modelos AR (p), MA (q), ARMA (p,q) Y ARIMA(p,d,q))

El estudio y análisis de una serie de tiempo bajo el enfoque del dominio del tiempo consiste en estudiar sus autocorrelaciones, es decir analizar la relación que existe entre los datos y que tan relacionados están, con el fin de identificar el proceso que genero a la serie, de este análisis podemos reconocer los modelos que caracterizan a nuestra serie según el comportamiento de sus autocorrelaciones. A continuación se describen los modelos usados en el enfoque del dominio del tiempo:

II.2.1 Modelos Autoregresivos AR(p)

Son procesos estocásticos lineales y estacionarios, el término autoregresivo se debe a que este proceso se obtiene de una combinación lineal de valores desfasados de la misma variable.

En particular un modelo AR (1) se representa de la siguiente manera:

$$y_t = \phi_1 y_{t-1} + a_t \quad (II.1)$$

Usando el operador de retardo B se tiene:

$$(1 - \phi_1 B) y_t = a_t \quad (II.2)$$

Condición de estacionariedad

Para que un modelo AR (1) sea estacionario las raíces de la ecuación $1 - \phi_1 B = 0$ deben caer fuera del círculo unitario, entonces:

$$\left| \frac{1}{\phi_1} \right| = |B| > 1 \quad |\phi_1| < 1$$

Todo modelo autoregresivo es invertible.

Función de Autocovarianza

$$\gamma_k = \begin{cases} \gamma_0 = \frac{\sigma_a^2}{1 - \phi_1^2} & \mathbf{k} = 0 \\ \gamma_k = \phi_1 \gamma_{k-1} & \mathbf{k} > 0 \end{cases} \quad (II.3)$$

Función de Autocorrelación

$$\rho_k = \begin{cases} \rho_0 = 1 & \mathbf{k} = 0 \\ \rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k & \mathbf{k} > 0 \end{cases} \quad (II.4)$$

En general un modelo AR (p) se representa de la siguiente manera:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t \quad (II.5)$$

Condición de Estacionariedad

Para que un modelo AR (1) sea estacionario las raíces de la ecuación:

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0 \quad (11.6)$$

deben caer fuera del círculo unitario, todo modelo autoregresivo es invertible.

Función de Autocovarianza

$$\gamma_k = \begin{cases} \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_a^2 & \mathbf{k} = 0 \\ \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} & \mathbf{k} > 0 \end{cases} \quad (11.7)$$

Función de Autocorrelación

$$\rho_k = \begin{cases} \rho_0 = 1 & \mathbf{k} = 0 \\ \rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} & \mathbf{k} > 0 \end{cases} \quad (11.8)$$

II.2.2 Modelos de Medias Móviles MA(q)

Son procesos estocásticos lineales y estacionarios, el término media se refiere a que este proceso se obtiene mediante una combinación lineal de variables aleatorias de la forma de ruido blanco, el término móviles se refiere a que estas medias varían a lo largo del tiempo.

En particular un modelo MA (1) se representa de la siguiente manera:

$$y_t = a_t + \theta_1 a_{t-1}$$

Usando el operador de retardo B se tiene:

$$y_t = (1 - \theta_1 B) a_t$$

Condición de invertibilidad

Para que un modelo MA (1) sea invertible las raíces de la ecuación $1 - \theta_1 B = 0$ deben caer fuera del círculo unitario, entonces:

$$\left| \frac{1}{\theta_1} \right| = |B| > 1 \qquad |\theta_1| < 1$$

Todo modelo de medias móviles es estacionario.

Función de Autocovarianza

$$\gamma_k = \begin{cases} \gamma_0 = (1 + \theta_1^2) \sigma_a^2 & \mathbf{k} = 0 \\ \gamma_1 = -\theta_1 \sigma_a^2 & \mathbf{k} = 1 \\ \gamma_k = 0 & \mathbf{k} > 1 \end{cases} \quad (\text{II.9})$$

Función de Autocorrelación

$$\rho_k = \begin{cases} \rho_0 = 1 & \mathbf{k} = 0 \\ \rho_1 = \frac{-\theta_1}{1 + \theta_1^2} & \mathbf{k} = 1 \\ \rho_k = 0 & \mathbf{k} > 1 \end{cases} \quad (\text{II.10})$$

En general un modelo MA (q) se representa de la siguiente manera:

$$y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (\text{II.11})$$

Condición de invertibilidad

Para que un modelo MA (q) sea invertible las raíces de la ecuación:

$$1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0 \quad (\text{II.12})$$

deben caer fuera del círculo unitario, todo modelo media móvil es estacionario.

Función de Autocovarianza

$$\gamma_k = \begin{cases} \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_a^2 & \mathbf{k} = 0 \\ \gamma_k = (-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma_a^2 & \mathbf{k} = 1, 2, \dots, q \\ \gamma_k = 0 & \mathbf{k} > q \end{cases} \quad (\text{II.13})$$

Función de Autocorrelación

$$\rho_k = \begin{cases} \rho_0 = 1 & \mathbf{k} = 0 \\ \rho_k = \frac{(-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q)}{(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)} & \mathbf{k} = 1, 2, \dots, q \\ \rho_k = 0 & \mathbf{k} > q \end{cases} \quad (\text{II.14})$$

II.2.3 Modelos ARMA (p,q)

Son procesos estocásticos lineales y estacionarios, autoregresivos y de medias móviles, se representan de la siguiente manera:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (\text{II.15})$$

Utilizando el operador de retardo B tenemos:

$$\phi(B)y_t = \theta(B)a_t$$

Para que el proceso sea estacionario las raíces de la ecuación:

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0 \quad (\text{II.16})$$

Deben caer fuera del círculo unitario y el modelo tiene representación MA (∞) de la siguiente manera:

$$y_t = \frac{\theta(B)}{\phi(B)} a_t = \psi(B) a_t \quad (II.17)$$

Donde:

$$\phi(B) \psi(B) = \theta(B)$$

Para que el proceso sea invertible las raíces de la ecuación:

$$1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0 \quad (II.18)$$

Deben caer fuera del círculo unitario y el modelo tiene representación $AR(\infty)$ de la siguiente manera:

$$a_t = \frac{\phi(B)}{\theta(B)} y_t = \pi(B) y_t \quad (II.19)$$

Donde:

$$\theta(B) \pi(B) = \phi(B)$$

II.3 Integración y Diferenciación

La integración y la diferenciación utilizada en los modelos ARIMA, con el fin de conseguir estacionariedad (diferenciando) y en el caso de la integración para volver a la serie original una vez diferenciada, son aplicadas a través de sus operadores, estos operadores se definen de la siguiente manera:

Operador de diferenciación

$$D = \nabla^d = (1 - B)^d$$

B: Es llamado el Operador de retardo

Ejemplo: Si $d = 1$ entonces $\nabla y_t = (1 - B)y_t = y_t - y_{t-1} = u_t$

donde u_t es un proceso estacionario, es decir u_t se obtiene diferenciando d veces y_t .

Operador de Integración

$$S = \nabla^{-d} = (1 - B)^{-d}$$

B: Operador de retardo

Ejemplo: Si $d=1$ entonces $y_t = (1 - B)^{-1} u_t = \nabla^{-1} u_t$

y_t se obtiene integrando d veces u_t , es decir y_t es un proceso integrado de orden d .

Nótese que en ambos casos el d (parámetro de diferenciación) se asume como entero.

II.4 Modelos ARIMA (p,d,q)

Los modelos ARIMA constituyen una clase particular de procesos no estacionarios. En la realidad buena parte de las series económicas se deben considerar generadas por procesos no estacionarios.

A un proceso integrado y_t se le denomina proceso ARIMA (p,d,q) si tomando primeras diferencias d veces se obtiene un proceso estacionario u_t del tipo ARMA (p,q). La "I" central del término ARIMA indica integrado. Así pues se tendría que:

$$u_t = \nabla^d y_t = (1 - B)^d y_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) u_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (\text{II.20})$$

O substituyendo $u_t = (1 - B)^d y_t$ en la ecuación (II.20):

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) (1 - B)^d y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (\text{II.21})$$

En forma más compacta, un proceso ARIMA (p,d,q) se define así:

$$\phi(B)(1-B)^d y_t = \theta(B)a_t$$

II.5 Análisis bajo el enfoque de dominio de frecuencia

En lo que va el desarrollo de la investigación, hemos repasado la teoría concerniente en el dominio del tiempo donde

las autocorrelaciones juegan un papel importante, sin embargo una serie no solo se puede analizar por sus autocorrelaciones, también podemos analizar una serie identificando cuales son las frecuencias que mejor explican su comportamiento y si es que tiene periodicidades, poder descubrir con qué frecuencia se dan, todo esto se analiza bajo el enfoque del dominio de frecuencia y para ello es necesario conocer una serie de conceptos y herramientas que se desarrollan a continuación:

II.5.1 Función Periódica

Una función $f(x)$ se dice periódica p si para todo k entero:

$$f(x + kp) = f(x)$$

Es decir, son funciones cuyos valores se repiten después de un cierto intervalo de valores de x a cuya longitud llamamos periodo.

II.5.2 Cosenoidales y Sinusoidales

Son casos particulares de funciones periódicas que se generan a partir de funciones coseno y seno respectivamente, las funciones $f(t) = A \cos wt$ y $f(t) = B \cos wt$ son funciones periódicas con periodo $2\pi / w$ ya que:

$$\begin{aligned} f(t + 2\pi / w) &= A \cos w(t + 2\pi / w) \\ &= A \cos(wt + 2\pi) \\ &= A \cos wt = f(t) \end{aligned}$$

A la constante “A” se le llama amplitud y determina los valores máximo y mínimo (A y -A respectivamente) que toma la función $f(t)$

El periodo es $2\pi/\omega$, es decir hay que dar un valor a ω tal que cada $2\pi/\omega$ unidades de tiempo se repita el valor de $f(t)$.

Si la función demora $2\pi/\omega$ unidades de tiempo en volver a tomar el mismo valor, es decir en completar un ciclo, eso significa que en una unidad de tiempo se completará una porción $1/(2\pi/\omega)$ del ciclo es decir $\omega/2\pi$.

A $\omega/2\pi$ se le denomina "frecuencia" y se le denota con $f = \omega/2\pi$.

A $\omega = f \cdot 2\pi$ se le denomina "frecuencia angular".

II.5.3 Armónica

La suma de cosenoidales y sinusoidales (que son funciones periódicas) del mismo periodo (aunque pueden tener diferentes amplitudes) se le denomina una armónica.

$$A \cos \omega t + B \sin \omega t$$

Y tiene las siguientes características:

-Es periódica con periodo $2\pi/\omega$, cada $2\pi/\omega$ unidades de tiempo se repite el valor de la función, por ejemplo si $\omega = \pi/6$ significa que cada 12 unidades de tiempo se repite el valor de la función.

-Sus valores máximo y mínimo son C y $-C$ respectivamente donde $C = \sqrt{A^2 + B^2}$ recibe el nombre de amplitud de la armónica.

-La frecuencia es $\omega/2\pi$, la frecuencia angular es ω , significa que en una unidad de tiempo se completa una porción $\omega/2\pi$ del ciclo, por ejemplo si $\omega = \pi/6$ significa que en una unidad de tiempo se completa la doceava parte de un ciclo.

II.5.4 Análisis de Fourier y análisis armónico

Al representar una serie de tiempo como una suma finita de armónicas estamos haciendo un análisis armónico, el antecedente de este análisis es el análisis de Fourier.

El análisis de Fourier trata básicamente de la aproximación de funciones mediante la suma infinita de funciones del tipo de las armónicas.

Existe un teorema que dice que toda función que sea continua por trozos y cuya derivada sea también continua por trozos puede ser aproximada mediante una suma infinita de senos y cosenos.

La serie de Fourier para aproximar la función $f : [-L, L] \rightarrow R$ se define como:

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi}{L}x + b_n \operatorname{sen} \frac{n\pi}{L}x$$

Donde

$$b_n = \frac{1}{L} \int_{-L}^L f(x) \operatorname{sen} \frac{n\pi}{L}x dx \quad n = 1, 2, 3, \dots \quad (\text{II.22})$$

$$a_n = \frac{1}{L} \int_{-L}^L f(x) \cos \frac{n\pi}{L}x dx \quad n = 0, 1, 2, 3, \dots \quad (\text{II.23})$$

Si en esa definición se reemplazan sen y cos por su expresión en términos de e, utilizando las formulas de Euler:

$$\cos \alpha = \frac{e^{i\alpha} + e^{-i\alpha}}{2} \quad \operatorname{sen} \alpha = \frac{e^{i\alpha} - e^{-i\alpha}}{2}$$

se obtiene lo que se conoce como la forma compleja de la serie de Fourier:

$$\sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi}{L}x}$$

Donde

$$c_n = \frac{1}{L} \int_{-L}^L f(x) e^{\frac{in\pi}{L}x} dx$$

El análisis armónico pretende representar el comportamiento de la serie como una suma finita de armónicas más perturbaciones puramente aleatorias.

II.5.5 Periodograma de Schuster

El periodograma tiene como principal objetivo, confrontar todas las frecuencias e identificar cual es la que mejor explica el comportamiento de la serie, se define de la siguiente manera:

Para $w = \frac{2\pi p}{N}$ $p = 1, \dots, \left(\frac{N}{2}\right) - 1$

Se calculan

$$a(w) = \frac{2}{N} \sum_t y_t \cos tw \quad b(w) = \frac{2}{N} \sum_t y_t \sin tw$$

En base a $b(w)$ y $a(w)$ se calcula la intensidad de cada frecuencia w como:

$$I(w) = [a(w)]^2 + [b(w)]^2$$

Al gráfico de $I(w)$ vs. w se le denomina periodograma.

II.5.6 Análisis espectral de un proceso estacionario

Sea y_t un proceso estacionario con autocovarianzas γ_k , absolutamente convergentes, esto es, $\sum |\gamma_k| < \infty$. La función espectral de y_t esta dada por:

$$f(w) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i w k} \quad w \in [-\pi, \pi] \quad (\text{II.24})$$

$$f(w) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(wk) \right] \quad (\text{II.25})$$

Donde se consideran las propiedades $\gamma_k = \gamma_{-k}$, $\text{sen}(-wk) = -\text{sen}(wk)$ y $\cos(-wk) = \cos(wk)$.

Propiedades de $f(w)$:

$f(w)$ es una función continua y real.

$f(w) = f(-w)$ para todo w .

$f(w) \geq 0$ para todo w .

II.5.7 Función espectral de un modelo ARMA:

Sea y_t un proceso ARMA (p, q) de la forma $\phi(B)(y_t - \mu) = \theta(B)a_t$ y a_t un proceso de ruido blanco con media 0 y varianza constante σ_a^2 . La función espectral de y_t esta dada por:

$$f(w) = \frac{\sigma_a^2}{2\pi} \left| \frac{\theta_q(e^{-iw})}{\phi_p(e^{-iw})} \right|^2 \quad (\text{II.26})$$

Donde $\phi_p(\cdot)$ y $\theta_q(\cdot)$ son polinomios del proceso ARMA (p, q). (Harvey (1981)).

II.6 Modelos ARFIMA

Los modelos ARFIMA constituyen una buena alternativa para modelar series que presenta memoria larga, ya que al permitir que el parámetro de diferenciación sea fraccional, la serie la cual estemos analizando no perderá el efecto de memoria larga. Los usos más importantes de los modelos ARFIMA son los siguientes:

Se pueden modelar comportamientos que permanecen en periodos largos (persistentes).

Se puede incorporar un número grande de términos autoregresivos y media móvil para poder capturar información contenida en los datos.

Flexibilidad para explicar estructuras de correlación en el corto y largo plazo de la serie.

II.6.1 Definición y propiedades

Un proceso y_t es integrado de orden d ($y_t \sim I(d)$) si:

$$(1 - B)^d y_t = u_t \quad (II.27)$$

donde u_t es un proceso estacionario con densidad espectral acotada en la frecuencia cero. Cuando $d = 0$ el proceso y_t es un proceso estacionario, cuando $d = 1$ el proceso y_t es integrado de orden 1 ($I(1)$) con una raíz unitaria, cuando d es un número entero mayor que 1 se dice el proceso y_t es un proceso integrado de orden d ($I(d)$) con d raíces unitarias. Cuando d es un número real no entero se dice que el proceso y_t es un proceso fraccionalmente integrado, si además el proceso u_t es un proceso ARMA (p,q) de la forma $\phi(B)u_t = \theta(B)a_t$ entonces y_t es un proceso ARMA (p,q) fraccionalmente integrado o ARFIMA (p,d,q) .

Cuando $u_t = a_t$ se obtiene el más sencillo de los procesos fraccionalmente integrados ARFIMA (0, d, 0) donde a_t es ruido blanco con media cero y varianza constante σ_a^2 entonces el proceso y_t se expresa de la siguiente manera:

$$(1 - B)^d y_t = a_t \quad (II.28)$$

Donde B es el operador de retardo y el polinomio $(1 - B)^d$ se define para cualquier $d > -1$ por la expansión binomial:

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{d(1-d)B^2}{2!} - \frac{d(1-d)(2-d)B^3}{3!} - \dots \quad (II.29)$$

Condición de Estacionariedad:

Al igual que en los modelos ARIMA en los modelos ARFIMA se cumple con la condición de estacionariedad pero solo para $d < 0.5$ el proceso ARFIMA (0,d,0) es estacionario y se puede representar como un $MA(\infty)$ de la siguiente manera:

$$y_t = (1 - B)^{-d} a = \sum_{k=0}^{\infty} \psi_k a_{t-k} \quad (II.30)$$

Con
$$\psi_k = \frac{\Gamma(k + d)}{\Gamma(d)\Gamma(k + 1)}$$

Condición de Invertibilidad:

Para $d > -0.5$ el proceso ARFIMA (0, d, 0) es invertible y se puede representar como un $AR(\infty)$ de la siguiente manera:

$$(1 - B)^d y_t = \sum_{k=0}^{\infty} \pi_k y_{t-k} = a_t \quad (II.31)$$

Con
$$\pi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

La función de densidad espectral viene dada por:

$$f(\lambda) = \frac{\sigma_a^2}{2\pi} |2\text{sen}(\lambda/2)|^{-2d} \quad (\text{II.32})$$

Cuando $\lambda \rightarrow 0$ entonces $f(\lambda) = \frac{\sigma_a^2}{2\pi} \lambda^{-2d}$ de aquí se deduce que para valores $0 < d < 1/2$ la función de densidad espectral es una función decreciente de λ no acotada en el origen y por lo tanto presenta memoria larga.

La varianza y autocovarianza vienen dados de la siguiente manera:

$$\gamma(0) = \sigma_a^2 \frac{\Gamma(1-2d)}{(\Gamma(1-d))^2} \quad (\text{II.33})$$

$$\gamma(k) = \sigma_a^2 \frac{\Gamma(1-2d)\Gamma(k+d)}{\Gamma(d)\Gamma(1-d)\Gamma(k+1-d)}, \quad \text{para } k \geq 1 \quad (\text{II.34})$$

La función de autocorrelación viene dada de la siguiente manera:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} = \prod_{1 \leq i \leq k} \frac{i+d-1}{i-d}, \quad k \geq 1 \quad (\text{II.35})$$

Si $k \rightarrow \infty$ y aplicando $(\Gamma(a+x)/\Gamma(b+x) = x^{a-b})$ entonces:

$$\rho(k) = \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \quad (\text{II.36})$$

Las principales características de los modelos ARFIMA (0,d,0) son las siguientes: para valores $0 < d < 1/2$ la función densidad espectral es una función decreciente de λ no acotada en el origen y concentrada en bajas frecuencias, sus autocorrelaciones decrecen lentamente de forma hiperbólica a razón k^{2d-1} , son

positivas y no son sumables; estas características son propias de procesos con memoria larga.

Ana Pérez Espartero [10] en su trabajo muestra el siguiente ejemplo ilustrativo que diferencia a los modelos ARIMA (0,d,0) de los ARFIMA(0,d,0) “Para visualizar esta dependencia a largo plazo típica de los modelos ARFIMA el siguiente cuadro compara las autocorrelaciones de un proceso ARFIMA (0,1/3,0) con las de un proceso AR(1) de la forma $y_t = \phi y_{t-1} + a_t$, con $\phi = 0.5$. Aunque en ambos casos la autocorrelacion de orden uno es igual a 0.5, se aprecia claramente una gran diferencia en el ritmo al que las correlaciones convergen hacia cero, mientras que en el proceso AR (1) la autocorrelación de orden 25 es 2.98×10^{-8} , en el proceso fraccionalmente integrado es 0.173 y sólo se reduce a 0.109 en el retardo cien.”

Retardo (K)	$\rho(K)$ [ARFIMA (0,1/3,0)]	$\rho(K)$ [AR (1), con $\phi = 0.5$]
1	0,500	0,500
2	0,400	0,250
3	0,350	0,125
4	0,318	0,063
5	0,295	0,031
10	0,235	0,001
25	0,173	$2,98 \times 10^{-8}$
50	0,137	$8,88 \times 10^{-16}$
100	0,109	$7,89 \times 10^{-31}$

Tabla II.1

Otra de las características que diferencia claramente los modelos ARFIMA de los modelos ARMA de corta memoria y de los modelos integrados ARIMA es la forma como el output responde a un cambio unitario en la perturbación del modelo, diferente a los modelos ARIMA en los cuales el efecto de un shock sobre el nivel futuro de la serie es persistente y no hay reversión a la media, en los modelos ARFIMA el efecto acaba desapareciendo (aunque a un ritmo más lento que el exponencial de los modelos ARMA) y termina revirtiendo a la media.

Deibold y Rudebusch (1989) y Haubrich y Lo (1989) con la finalidad de medir el grado de persistencia de los shocks sobre la media, proponen utilizar la función respuesta al impulso acumulativa. Ana Pérez Espartero [10] también presenta en su trabajo la siguiente definición relacionada a la función respuesta al impulso acumulativa: “Dada una serie y_t , los valores de esta función son los coeficientes

del polinomio $c(B) = 1 + c_1 B + c_2 B^2 + \dots$, tal que $y_t = \zeta + c(B)a_t$. Así, el coeficiente c_k medirá el impacto que tiene una innovación unitaria en t sobre el nivel de la serie en $t+k$, y su límite $c_\infty = \lim_{k \rightarrow \infty} c_k$, medirá el impacto a largo plazo de dicha innovación. Cuando la serie y_t en primeras diferencias admite una representación de la forma:

$$(1 - B)y_t = a(B)a_t \quad (II.37)$$

el valor límite de la función respuesta al impulso acumulativa coincide con la suma de los coeficientes del polinomio $a(B)$. En concreto, si $a(B) = 1 + a_1 B + a_2 B^2 + \dots$, se comprueba fácilmente que $c_k = 1 + a_1 + \dots + a_k$, y por tanto el impacto a largo plazo que tiene una innovación unitaria sobre el nivel de la serie será:

$$c_\infty = \lim_{k \rightarrow \infty} c_k = \lim_{k \rightarrow \infty} (1 + a_1 + \dots + a_k) = a(1) \quad (II.38)$$

donde $a(1)$ es la suma de los coeficientes del polinomio $a(B)$.

En el modelo ARFIMA (p, d, q) , la forma del polinomio $a(B)$ se deduce inmediatamente de la ecuación $\phi(B)(1 - B)^d y_t = \theta(B)a_t$, obteniéndose:

$$a(B) = (1 - B)^{1-d} \phi(B)^{-1} \theta(B) \quad (II.39)$$

donde $(1 - B)^{1-d}$ admite una expansión binómica. Utilizando los resultados de Gradshteyn y Ryzhnik (1980) sobre la función hipergeométrica resulta que si $0 < d \leq 1$ $(1 - B)^{1-d} = F(d - 1, 1, 1; B)$, siendo F la función hipergeométrica. Esta función toma el valor $F(d - 1, 1, 1; 1) = 0$ si $d < 1$, y $F(d - 1, 1, 1; 1) = 1$ si $d = 1$, y por tanto:

$$a(1) = F(d - 1, 1, 1; 1) \phi(1)^{-1} \theta(1) = \begin{cases} 0 & \text{si } 0 < d < 1 \\ \phi(1)^{-1} \theta(1) & \text{si } d = 1 \end{cases} \quad (II.40)$$

De aquí se deduce que el efecto de un shock persiste indefinidamente en los procesos integrados ($d=1$) ya que $a(1) \neq 0$ y no hay reversión a la media, por el contrario en los procesos fraccionalmente integrados $I(d)$ con $0 < d < 1$, el efecto termina desapareciendo ($a(1)=0$) y la serie revierte finalmente a su media, aunque lo hará a un ritmo muy lento, que es caracterizado por el valor del parámetro d y no solamente se cumple en el caso estacionario ($0 < d < 1/2$), también se cumple en el caso en que el proceso es no estacionario ($1/2 < d < 1$), a consecuencia de ello suele decirse que un proceso ARFIMA no estacionario es “menos no estacionario” que los procesos con raíces unitarias.

Generalizando del modelo ARFIMA (0,d,0) si u_t es un proceso autoregresivo y de medias móviles ARMA(p,q) entonces y_t es un proceso ARMA fraccionalmente integrado o ARFIMA(p,d,q) de la forma:

$$\phi(B)(1-B)^d y_t = \theta(B)a_t \quad (II.41)$$

Donde $\phi(B)$ y $\theta(B)$ representan la parte autoregresiva y media móvil respectivamente del modelo en función del operador de retardo B , el polinomio $(1-B)^d$ se expresa de la misma forma que el de un ARFIMA (0,d,0) y a_t es el ruido blanco con media cero y varianza constante σ_a^2 .

Para valores de $d < 0.5$ y si todas las raíces de $\phi(B) = 0$ están fuera del círculo unitario el proceso y_t es estacionario y si $d > 0.5$ y todas las raíces de $\theta(B) = 0$ están fuera del círculo unitario el proceso y_t es invertible.

Para valores de $d < 0.5$ la función de densidad espectral del proceso ARFIMA (p,d,q) se define de la siguiente manera:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(e^{i\lambda})}{\phi(e^{i\lambda})} \right|^2 |4\sin^2(\lambda/2)|^{-d} \quad (II.42)$$

Cuando $\lambda \rightarrow 0$ resulta que $f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(1)}{\phi(1)} \right|^2 \lambda^{-2d}$ lo cual quiere decir que la función de densidad espectral no está acotada en el origen y por lo tanto el proceso y_t presenta memoria larga.

La garantía que para $-0.5 < d < 0.5$ el modelo es estacionario e invertible la da el siguiente teorema:

TEOREMA

Dado $\{y_t\}$ un proceso ARFIMA (p,d,q) entonces:

1) $\{y_t\}$ es estacionario si $d < 0.5$ y todas las raíces del polinomio $\phi(B) = 0$ caen fuera del círculo unitario.

2) $\{y_t\}$ es invertible si $d > -0.5$ y todas las raíces del polinomio $\theta(B) = 0$ caen fuera del círculo unitario.

Si $\{y_t\}$ es estacionario e invertible con densidad espectral $f(\lambda)$ y función de autocorrelación ρ_k entonces:

3) $\lim \lambda^{2d} f(\lambda)$ existe, cuando λ tiende a cero y es finito.

4) $\lim k^{1-2d} \rho(k)$ existe, cuando k tiende a cero y es finito.

II.6.2 La función gamma y los procesos de memoria larga

Se define la función gamma por:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx \quad (\text{II.43})$$

Para $p > 0$. Integrando por partes se obtiene:

$$\Gamma(p) = \left[-x^{p-1} e^{-x} \right]_0^{\infty} + (p-1) \int_0^{\infty} x^{p-2} e^{-x} dx = (p-1) \Gamma(p-1) \quad (\text{II.44})$$

Como $\Gamma(1) = 1$ la función gamma coincide con la función factorial para p entero. Para cualquier valor positivo y real de p esta función generaliza la función factorial, ya que siempre podemos escribir $p = n + r$ siendo n entero:

$$\Gamma(n+r) = (n+r-1)(n+r-2)\dots(1+r)\Gamma(1+r). \quad (\text{II.45})$$

Esta propiedad hace que la función gamma se tabule entre 1 y 2 porque podemos calcularla para cualquier p real aplicando esta relación y solo necesitamos el valor de $\Gamma(1+r)$, que corresponderá a un argumento de la función entre 1 y 2. Su valor en ese intervalo es siempre próxima a la unidad, y la función toma el valor mínimo para $1+r = 1.4616$, donde entonces la función vale $\Gamma(1.4616) = 0.8856$.

La función gamma puede también definirse para p negativo, mediante:

$$\Gamma(p+1) = \lim_{n \rightarrow \infty} \frac{n!}{(p+1)(p+2)\dots(p+n)} \quad (\text{II.46})$$

y se demuestra que este límite existe y que para valores positivos de volvemos al caso anterior y el límite coincide con la integral de definición para $p > 0$. Además, la función así definida sigue cumpliendo la ley recurrente $\Gamma(p) = (p-1)\Gamma(p-1)$. Se demuestra también que esta función se hace infinita para cero y los valores enteros negativos del argumento, $p = 0, -1, -2, -3$.

Estos resultados llevan a definir el operador diferencia fraccional por la suma infinita:

$$(1-B)^d = \sum_{i=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i B^i \quad (II.47)$$

si d es entero cuando $i > d$ entonces $\Gamma(d-i+1)$ corresponde a la gamma con argumento entero negativo que, según hemos visto, es infinita. Por tanto, todos los coeficientes serán cero para $i > d$ y obtendremos la expresión de Newton. Observemos que, como:

$$\frac{\Gamma(d+1)}{\Gamma(d-i+1)} = d(d-1)(d-2)\dots(d-i+1). \quad (II.48)$$

e igualmente

$$\frac{\Gamma(i-d)}{\Gamma(-d)} = (i-d-1)(i-d-2)\dots(1-d)(-d) \quad (II.49)$$

podemos escribir los términos del desarrollo general de $(1-B)^d$ mediante:

$$\frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i = \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}, \quad (II.50)$$

que es la forma habitual de expresar los coeficientes del desarrollo cuando p no es entero.

II.6.3 Identificación de modelos ARFIMA

Cuando trabajamos en el contexto de los modelos de Box y Jenkins, la identificación de los parámetros autoregresivos y medias móviles se basan en la función de autocorrelación estimada para identificar el proceso que genere la Serie de Tiempo estudiada.

Teniendo en cuenta que la teoría de los procesos ARMA (p, q) se desarrolla solo para procesos estocásticos estacionarios, el primer requisito que se debe tener en cuenta en la etapa de identificación es verificar si la Serie de Tiempo es

estacionaria o no, con este fin, podemos utilizar métodos formales tales como; los contrastes de raíces unitarias y el análisis exploratorio de la serie.

Por ejemplo, el gráfico de la serie mostrará si existe tendencia, cambios en el nivel, etc; el correlograma muestral ayudara a determinar si la serie tiene memoria larga (decrecimiento lento de las autocorrelaciones), un espectro estimado con valores muy elevados en frecuencias próximas a cero también son indicios de persistencia en periodos largos.

Para identificar memoria larga en series de tiempo existen varias dójimas para contrastar la hipótesis de memoria larga ($d > 0$) (que desarrollaremos mas adelante con más detalle). Otra herramienta que podemos incorporar para explorar si la serie presenta memoria larga es el gráfico de dispersión del logaritmo del periodograma versus el logaritmo de $(4\text{sen}^2(\frac{w_j}{2}))$ donde w_j es la frecuencia del periodograma, si estos puntos muestran una relación aproximadamente lineal indicará persistencia en periodos largos.

II.6.4 Dósimas para contrastar la hipótesis de memoria larga

Como hemos visto anteriormente en los conceptos previos, el termino de “memoria larga”, es una característica que poseen algunas series de tiempo según como se comportan sus autocorrelaciones, esto nos lleva a utilizar métodos para identificar la presencia de esta característica, pues sabemos que es el paso inicial antes de analizar una serie mediante el método ARFIMA, para evaluar si una serie de tiempo presenta o no “memoria larga”, existen varias pruebas que contrastan la hipótesis nula de “memoria corta”, contra la alternativa de “memoria larga”, algunas pruebas se describen a continuación:

La prueba de Geweke y Porter - Hudak (1983)

Este contraste semiparamétrico de larga memoria, se basa en la inferencia sobre el estimador del parámetro de diferenciación fraccional en la ecuación de regresión de mínimos cuadrados en el dominio de frecuencia

$$\log l_y(w_j) = c + d \log(4 \sin^2(w_j/2)) + e_j \quad (\text{II.51})$$

Donde $l_y(w_j)$ es el periodograma de y_t en la frecuencia $w_j = 2\pi j/T$ con $j = 1, \dots, m = \lceil T^{1/2} \rceil$, donde $\lceil x \rceil$ denota el mayor entero que es menor o igual a x . Geweke & Porter - Hudak (1983) argumentan que cuando $-0.5 < d < 0$ y se cumplen algunas condiciones adicionales, entonces

$$GPH = \frac{\hat{d} - d}{\text{var}(\hat{d})^{1/2}} \xrightarrow{\text{dist}} N(0, 1) \quad (\text{II.52})$$

donde $\text{var}(\hat{d})^{1/2} = \pi^2 / 6 (\sum_{j=1}^m (X_j - \bar{X})^2)^{-1}$ con $X_j = \log(4 \sin^2(w_j/2))$.

El estimador GPH es simple de aplicar y robusto a no normalidad; sin embargo, presenta problemas de sesgo y eficiencia en presencia de componentes a corto plazo con raíces cercanas al círculo unitario del proceso y_t .

El contraste de Kwiatowski, Phillips, Schmidt y Shin (1992) (KPSS)

El contraste KPSS, es desarrollado para contrastar la hipótesis nula de estacionariedad ($d=0$), también puede emplearse para probar larga memoria (Lee & Schmidt 1996). Esta prueba está basada en la representación estructural $y_t = \delta_0 + \delta_1 t + r_t + e_t$, $r_t = r_{t-1} + u_t$ donde e_t es un proceso estacionario y $u_t \sim i.i.d(0, \sigma_u^2)$ La hipótesis nula de corta memoria del contraste es $H_0 : \sigma_u^2 = 0$ contra $H_1 : \sigma_u^2 > 0$, donde el estadístico de prueba tipo, multiplicadores de Lagrange (LM), está dado por

$$LM_{KPSS} = T^{-2} \sum_{t=1}^T S_t^2 / \hat{\sigma}_e^2 \quad (11.53)$$

donde $S_t = \sum_{i=1}^T e_i$ es la suma parcial de los residuales de la regresión de y_t sobre una constante y una tendencia lineal, \hat{e}_t y $\hat{\sigma}_e^2 = \sum_{t=1}^T \hat{e}_t^2 / T$ es la varianza de los residuales.

La prueba de Robinson (1994)

Es una prueba semiparamétrica bajo el enfoque del dominio de frecuencia a partir del enfoque de la densidad espectral de y_t , para obtener estimaciones consistentes del coeficiente de Hurts (1951), $H = d + \frac{1}{2}$ o equivalente a $d = H - \frac{1}{2}$. El estimador consistente del coeficiente de Hurts para contrastar $H_0 : H = \frac{1}{2}$ contra $H_1 : H > \frac{1}{2}$ (equivalente a $H_0 : d = 0$ contra $H_1 : d > 0$) está dado por

$$\hat{H}_{m,b} = 1 - \frac{\log(\hat{F}(b_{wm})/\hat{F}(w_m))}{2\log(b)} \quad (11.54)$$

El cual está basado en el estimador del periodograma promedio entre la frecuencia 0 y $w = 2\pi m/T$, la distribución asintótica del estimador es

desconocida. Una de las desventajas de la prueba es su pobre comportamiento en términos del sesgo y error cuadrático medio.

La prueba propuesta por Elkin Castaño, Karoll Gomez y Santiago Gallon (2008)

Esta prueba presentada por Castaño, Gómez y Gallon está basada en una aproximación autoregresiva de la componente a corto plazo del modelo ARFIMA. La prueba tiene el siguiente procedimiento:

Se considera el proceso ARFIMA (p,d,q) estacionario e invertible, $\phi(B)(1-B)^d y_t = \theta_0 + \theta(B)a_t$ donde los polinomios $\phi(B)$ y $\theta(B)$ satisfacen las mismas condiciones de regularidad exigidas para el modelo ARMA(p,q).

Si $d > -0.5$, es decir el proceso es invertible entonces se puede reescribir como

$$(1-B)^d \pi(B) y_t = a_t$$

donde $\pi(B) = \phi(B)/\theta(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$ es la componente a corto plazo del modelo ARFIMA.

Según Said y Dickey (1984), para cierto orden p^* suficientemente grande, el proceso anterior puede ser aproximado como

$$(1-B)^d \pi^*(B) y_t = a_t$$

Donde $\pi^*(B) = 1 - \pi_1^* B - \dots - \pi_p^* B^p$

Basados en este modelo aproximado $ARFIMA(p^*, d, 0)$ el procedimiento propuesto sugiere contrastar la hipótesis nula de memoria corta, $H_0 : d = 0$ con la alternativa de memoria larga $H_1 : d > 0$ empleando los resultados asintóticos de la distribución del estimador máximo verosímil del parámetro d . El estadístico de prueba es

$$t_d = \frac{d}{se(d)} \quad (II.55)$$

Donde d es el estimador máximo verosímil de d $se(d)$ es la estimación de su error estándar y bajo H_0 $t_d \underline{dist} N(0,1)$.

Las propiedades de tamaño y potencia dependen de la selección de p^* .

II.7 Métodos de estimación para el parámetro de diferenciación fraccional “d”

Existen varios métodos para estimar el orden de diferenciación “d”, el método de regresión utilizando la función de periodograma propuesto por Geweke y Porter - Hudak (1983), el método de regresión utilizando la función de periodograma suavizado propuesto por Reisen (1993, 1994), el método basado en coeficientes de Hurst (1951, 1956) y el método de Máxima Verosimilitud.

II.7.1 Método de regresión usando la función de periodograma

Este método utiliza la función de periodograma para luego expresarla como una ecuación de regresión, con el fin de obtener una estimación para “d” por el método de mínimos cuadrados.

Definición: Sea y_t un proceso ARIMA (p,d,q) $d(-0.5,0.5)$ representado por $(1-B)^d y_t = U_t$, donde $\phi(B)U_t = \theta(B)a_t$ y a_t es un proceso de ruido blanco con media 0 y varianza σ_a^2 .

Teorema: Sea y_t un proceso ARIMA (p,d,q) con $d(0,0.5)$, se tiene:

- a) $\sum \rho_k$ es infinito
- b) $f(w) \rightarrow \infty$ cuando $w \rightarrow 0$. Entonces, y_t es un proceso con propiedades de “memoria larga”. Para $d(-0.5,0)$, y_t muestra propiedades de “memoria corta”.

Por el teorema anterior la función de densidad espectral y_t es dada por:

$$f(w) = f_u(w)(2\text{sen}(w/2))^{-2d}, w(-p, p) \quad (II.56)$$

$f_u(w)$ es la función espectral de U_t . Aplicando logaritmo a $f(w)$ se tiene:

$$\ln f(w) = \ln f_u(w) - d \ln(2 \operatorname{sen}(w/2))^2 \quad (\text{II.57})$$

Escribiendo la expresión anterior de otra forma se tiene:

$$\ln f(w) = \ln f_u(0) - d \ln(2 \operatorname{sen}(w/2))^2 + \ln \left[\frac{f_u(w)}{f_u(0)} \right] \quad (\text{II.58})$$

Substituyendo w por $w_j = \frac{2\pi j}{n}$. Adicionando $\ln(I(w_j))$ entonces se tiene:

$$\ln I(w_j) = \ln f_u(0) - d \ln(2 \operatorname{sen}(w_j/2))^2 + \ln \left[\frac{f_u(w_j)}{f_u(0)} \right] + \ln \left[\frac{I(w_j)}{f(w_j)} \right] \quad (\text{II.59})$$

Se considera el límite superior de j igual a $g(n)$, el cual es escogido satisfaciendo

$\frac{g(n)}{n} \rightarrow 0$ cuando $n \rightarrow \infty$ y para w_j próximo a cero, $w_j \leq w_{g(n)}$ donde $w_{g(n)}$ es

pequeño. El término $\ln \left[\frac{f_u(w_j)}{f_u(0)} \right]$ es considerado despreciable comparado

con sus otros términos. Por lo tanto se obtiene una ecuación aproximada:

$$\ln I(w_j) \cong \ln f_u(0) - d \ln(2 \operatorname{sen}(w_j/2))^2 + \ln \left[\frac{I(w_j)}{f(w_j)} \right] \quad (\text{II.60})$$

Esta ecuación resulta una ecuación de regresión simple de la forma:

$$y_j = a + bx_j + e_j \quad j = 1, 2, \dots, g(n) \quad (\text{II.61})$$

Donde $y_j = \ln I(w_j)$ $x_j = \ln(2 \operatorname{sen}(w_j/2))^2$ $e_j = \ln \left[\frac{I(w_j)}{f(w_j)} \right] + c$ $b = -d$

$$a = \ln f_u(0) - c \text{ y } c = E \left(-\ln \left[\frac{I(w_j)}{f(w_j)} \right] \right).$$

Lema: Sea y_t un proceso ARIMA (p,d,q): $y_t = (1-B)^d \frac{\theta(B)}{\phi(B)} a_t$ con $d < 0$ y a_t

Un proceso de ruido blanco con distribución normal. Cuando $n \rightarrow \infty$ la secuencia

$-\ln \left[\frac{I_x(w_j)}{f(w_j)} \right]$, $w_j = \frac{2\pi j}{n}$ ($j=1,2,\dots, \left\lfloor \frac{n}{2} \right\rfloor$) distribuidos independientemente

con distribución Gumbel con media 0.577216 (constante de Euler) y varianza $\frac{\pi^2}{6}$.

Por el lema anterior cuando $d(-0.5,0)$ las variables $\ln \left[\frac{I(w_j)}{f(w_j)} \right]$, $j=1,2,\dots, g(n)$, son aproximadamente independientes con distribución Gumbel

con media -0.577216 y varianza $\frac{\pi^2}{6}$. Por lo tanto, la variable a_t también es

aproximadamente independiente con distribución Gumbel con media cero y varianza $\frac{\pi^2}{6}$.

Este resultado sugiere un estimador de d por el método de mínimos cuadrados de regresión de $y_1, y_2, \dots, y_{g(n)}$ en $x_1, x_2, \dots, x_{g(n)}$ donde g(n) es escogida tal que

cuando $n \rightarrow \infty$, $g(n) \rightarrow \infty$ y $\frac{g(n)}{n} \rightarrow 0$. Se tiene entonces un estimador:

$$\hat{b} = \frac{\sum_{i=1}^{g(n)} (x_i - \bar{x}) y_i}{\sum_{i=1}^{g(n)} (x_i - \bar{x})^2} \quad (II.62)$$

El estimador de d usando la función de periodograma en el método de regresión es dado por:

$$d_p = -\hat{b}$$

Con las propiedades:

$$E(d_p) = d \text{ y } \text{var}(d_p) = \frac{\pi^2}{6 \sum_{i=1}^{g(n)} (x_i - \bar{x})^2}$$

La distribución asintótica de d_p es dada por el siguiente teorema, el cual sugiere que $g(n) = n^\alpha$ ($0 < \alpha < 1$).

Teorema: (Geweke y Porter-Hudak (1983)). Sea y_t un proceso ARIMA (p,d,q) con $d < 0$ e $I(w_j)$ siendo la función de periodograma de y_t con frecuencias $w_j = \frac{2\pi j}{n}$.

Sea $d_p = -\hat{b}$ el estimador de d y suponiendo que $g(n)$ satisface $g(n) \rightarrow \infty$ y $\frac{g(n)}{n} \rightarrow 0$ cuando $n \rightarrow \infty$. Entonces $d_p = d$.

Sea $\lim |(\ln n)^2 / g(n)| = 0$ entonces:

$\frac{d_p - d}{\sqrt{\text{var}(d_p)}}$ está distribuido asintóticamente $N(0,1)$, donde $\text{var}(d_p)$ es la varianza de d_p .

II.7.2 Método de regresión usando la función de periodograma suavizado

Este método es similar al presentado anteriormente con la diferencia que utiliza la función del periodograma suavizado.

De la expresión (II.58) y usando los resultados asintóticos para $\ln \left[\frac{f_s(w)}{f(w)} \right]$, se puede escribir la ecuación de regresión de la forma:

$$\ln f(w) = \ln f_u(0) - d \ln(2 \sin(w_j/2))^2 + \ln \left[\frac{f_s(w_j)}{f_u(w_j)} \right] + \ln \left[\frac{f_u(w_j)}{f_u(0)} \right] \quad (\text{II.63})$$

Restringiendo el dominio de j , $1 \leq j \leq g(n)$ y escogiendo la función $g(n)$ de la forma anterior se puede escribir (II.63) de la siguiente forma:

$$\ln f(w) \cong \ln f_u(0) - d \ln(2 \sin(w_j/2))^2 + \ln \left[\frac{f_s(w_j)}{f(w)} \right] \quad (\text{II.64})$$

La ecuación (II.64) es una forma aproximada de la ecuación de regresión lineal simple (II.61) donde $y_j = \ln f_s(w_j)$ $x_j = \ln(2\sin(w_j/2))^2$

$$e_j = \ln \left[\frac{f_s(w_j)}{f(w_j)} \right] + c \quad b = -d \quad y \quad a = \ln f_u(0).$$

Como se a notado anteriormente, cuando $d(-0.5,0)$ en un modelo ARIMA (p,d,q) las variables e_j son asintóticamente no correlacionadas con media 0 y varianza dada de la forma:

$$\text{var} \left| \ln \left(\frac{f_s(w)}{f(w)} \right) \right| \approx \begin{cases} 0.539285 \left(\frac{m}{n} \right) & w \neq 0.\pi \\ 1.07856 \left(\frac{m}{n} \right) & w = 0.\pi \end{cases} \quad (\text{II.65})$$

esto sugiere un método de regresión para estimar d en la ecuación (II.61) donde $g(n)$ es escogido de la forma anterior es decir $g(n) = n^\alpha$ ($0 < \alpha < 1$).

El estimador de d obtenido por el método de regresión utilizando la función del periodograma suavizado es dado por:

$$d_{sp} = -\hat{b}$$

Donde

$$\hat{b} = \frac{\sum_{i=1}^{g(n)} (x_i - \bar{x}) y_i}{\sum_{i=1}^{g(n)} (x_i - \bar{x})^2} \quad (\text{II.66})$$

Con varianza

$$\text{var}(d_{sp}) \approx 0.53928 \frac{m}{n \sum_{i=1}^{g(n)} (x_i - \bar{x})^2}, w \neq 0.\pi \quad (\text{II.67})$$

(Obsérvese que $0.53928(m/n)$ es la varianza asintótica de e_j).

II.7.3 Método de Máxima Verosimilitud

Este método tiene una complejidad relacionada con la construcción de la función de verosimilitud y también en el proceso de evaluarla repetidamente; algunos autores como McLeod y Hipel (1978) y Brockwell y Davis (1987) argumentan estas desventajas, en consecuencia de estas desventajas otros autores como Brockwell y Davis (1987), Li y McLeod (1986) y Hosking (1984) sugieren usar una función aproximada de la función de verosimilitud; por otro lado Sowell (1990) plantea que las desventajas del método no tienen fundamento y él deriva la función exacta (no condicional) de verosimilitud para estimar d y lo compara empíricamente con los modelos propuestos por Fox y Taqqu (1986) y Geweke y Porter-Hudak (1983).

II.8 Procedimiento general para estimar los parámetros del modelo ARFIMA (p,d,q)

Sea y_t un modelo ARFIMA (p,d,q), entonces y_t puede ser escrito como $\phi(B)(1-B)^d y_t = \theta(B)a_t$, donde $\phi(B)$ y $\theta(B)$ son polinomios de orden p y q respectivamente, $\mu_x = 0$ y a_t es un proceso de ruido blanco con media cero y varianza 1. Se define $U_t = (1-B)^d y_t$, U_t es un modelo ARMA (p,q) y $z_t = \frac{\phi(B)}{\theta(B)} y_t$, z_t es un modelo ARFIMA(0,d,0).

1. Estimar d en el modelo ARFIMA (p,d,q), obteniendo \hat{d} .
2. Calcular $\hat{U}_t = (1-B)^{\hat{d}} y_t$.
3. Usar el procedimiento de Box-Jenkins para identificar y estimar los parámetros en el modelo ARMA (p,q), $\phi(B)\hat{U}_t = \theta(B)a_t$
4. Calcular $\hat{z}_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} y_t$.
5. Estimar d en el modelo ARFIMA (0, d, 0), $(1-B)^d \hat{z}_t = a_t$ el valor de \hat{d} es el nuevo estimador de d.

Repetir los pasos 2-5, hasta obtener una convergencia para d, ϕ y θ .

CAPITULO III

APLICACIÓN DE LA SERIE “IMPORTACION DE BIENES DE CONSUMO EN EL PERU (\$ MILLONES) “(1985-2008)

En toda economía de un país, las importaciones son el reflejo de cuanto la población demanda un bien o producto, en particular los bienes de consumo son todas aquellas mercancías producidas en el país o importadas para satisfacer directamente una necesidad como: alimentos, bebida, habitación, mobiliario, ornato, servicios personales, etc. Los bienes de consumo se pueden clasificar en dos grupos según sus características. Así tenemos:

- Bienes de consumo duradero
- Bienes de consumo no duradero

Estos bienes constituyen lo opuesto a los bienes de capital (como es el caso de las maquinarias), los cuales se utilizan para producir otros bienes.

Según el Banco Central de Reservas del Perú, en setiembre del 2009 las importaciones de bienes de consumo fueron de \$ 368 millones, mayor en 14.5 % al valor obtenido en el mes de agosto, en el mes de octubre fueron de \$ 372 millones mayores en 1 % al valor obtenido al mes de setiembre.

Sin embargo en el año 2008 las importaciones de bienes de consumo registraron su valor más alto con \$ 451 millones, en si en el periodo de los meses de setiembre – octubre del 2008 las importaciones de bienes de consumo sobrepasaron los \$ 400 millones, lo cual quiere decir, que en esa temporada la demanda de los bienes de consumo aumento considerablemente con respecto a otras.

Según la Superintendencia Nacional de Administración Tributaria (Sunat), hubo un descenso en las compras de bienes de consumo duradero (-18,5%) y bienes de consumo no duradero (-9,0%). Las importaciones de bienes de consumo no duradero

ascendieron en setiembre a US\$ 204 millones, registrando un descenso de 7,6% en comparación con similar mes del año 2008, al reducirse las importaciones de bebidas (-12,4%), otros bienes de consumo no duraderos (-11,9%), alimentos (-8,1%), tabaco (-6,9%), productos farmacéuticos y tocador (-4,2%) y vestuarios y otras confecciones (-3,9%). En relación a las importaciones de otros bienes no duraderos, su comportamiento negativo se explica por las menores compras de calzados, demás libros y folletos, los demás juguetes, entre otros.

Con respecto a los bienes de consumo duradero, a nivel desagregado, se observó un descenso en las importaciones de todos sus componentes tales como vehículos de transporte particular (-22,7%), objetos de adorno, instrumentos musicales y otros (-16,7%), muebles para el hogar (-16,0%), máquinas y aparatos de uso doméstico (-15,5%) y utensilios domésticos (-13,4%).

En el caso de los vehículos de transporte particular, y según origen, Japón fue el principal proveedor del mercado nacional con una participación del 46% del total, seguido de Corea del Sur (19%), China (11%) y Estados Unidos (6%).

Respecto a las menores importaciones de objetos de adorno, instrumentos musicales, esto se explicó principalmente por la reducción en las compras de los demás juegos activados con monedas, demás manufacturas de plásticos, bolsos de mano y artículos y material para cultura física y gimnasia o atletismo.

A continuación, se presenta el análisis de la serie de tiempo “Importación de bienes de consumo en el Perú (\$ mill)” (enero 1985-diciembre 2008) la cual tiene como fuente de información al Banco Central de Reserva del Perú (BCR), los valores del año 2009 los guardaremos para comparar los pronósticos resultantes.

III.1 Análisis descriptivo de la serie “Importación de Bienes de Consumo en el Perú (\$ mill)”

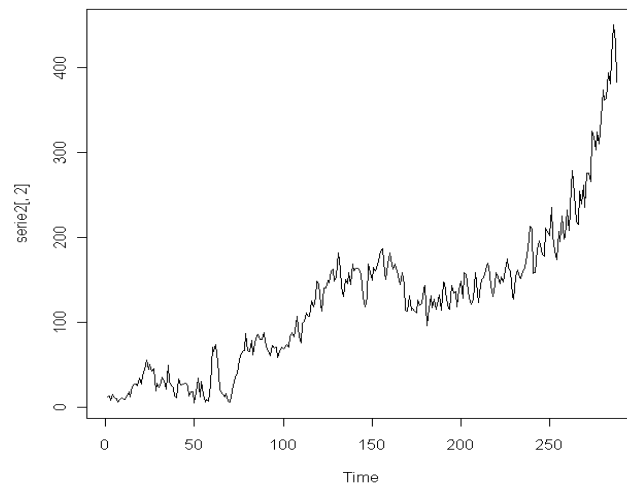


Grafico III.1

Observando el gráfico de la serie podemos ver que la serie es no estacionaria, presentando una tendencia creciente a lo largo del tiempo.

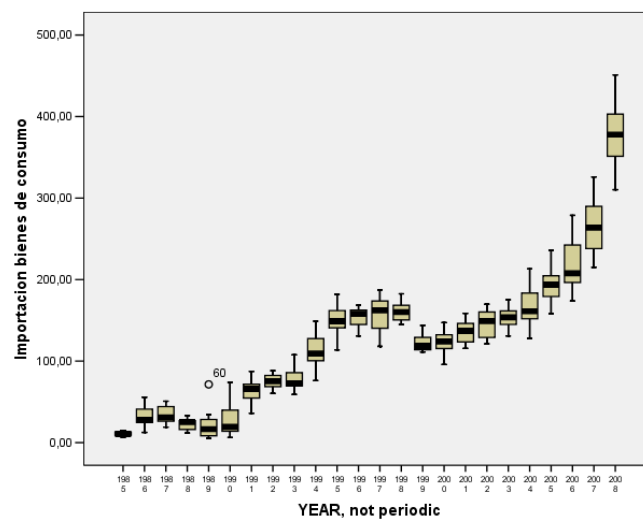


Grafico III.2

El gráfico de cajas simples corrobora lo antes visto en el gráfico de la serie, la tendencia es creciente y en los últimos años la variabilidad es creciente con respecto a cada año.

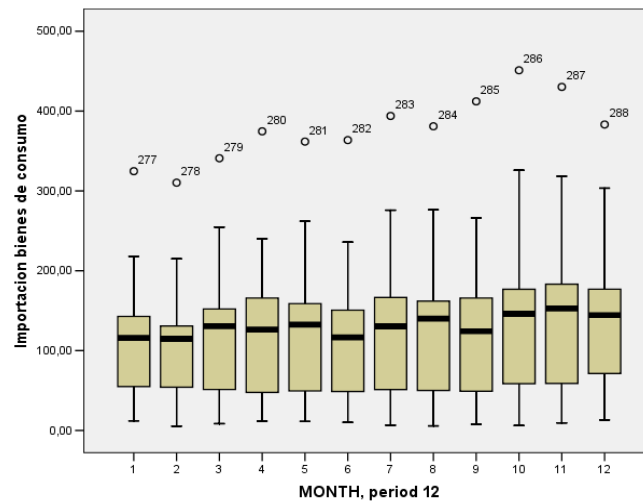


Grafico III.3

Observando el gráfico de cajas agregadas, la serie pareciera presentar estacionalidad pero esta no es marcada, por lo cual el análisis de la serie no presentará componente estacional, también se detectan datos atípicos para cada mes, estos datos son todos los del año 2008 ya que en ese año hubo un gran incremento de las importaciones de bienes de consumo con respecto a los otros años.

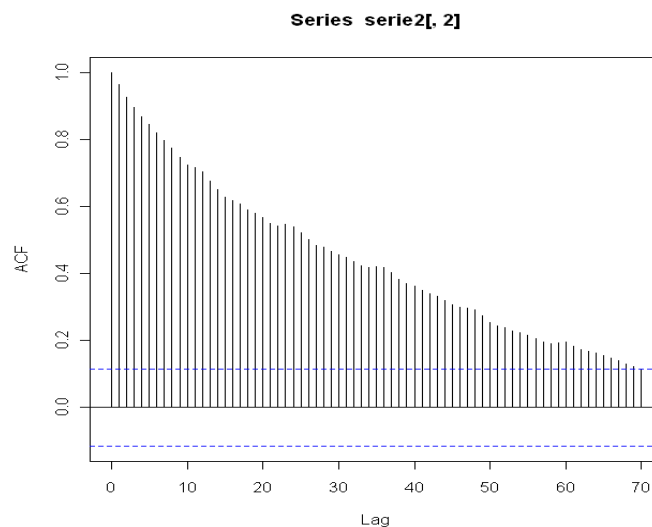


Grafico III.4

Observando éste gráfico se puede ver que las correlaciones decaen lentamente por lo que es un indicio de que la serie presentaría memoria larga, sin embargo esto se debe confirmar con otras herramientas.

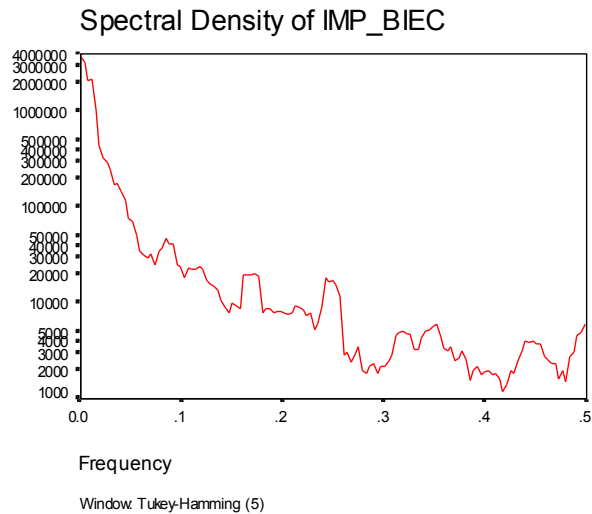


Grafico III.5

Observando éste gráfico vemos que la función se concentra para las bajas frecuencias, lo que indica que sus componentes a largo plazo deben tener un papel muy importante en el análisis.

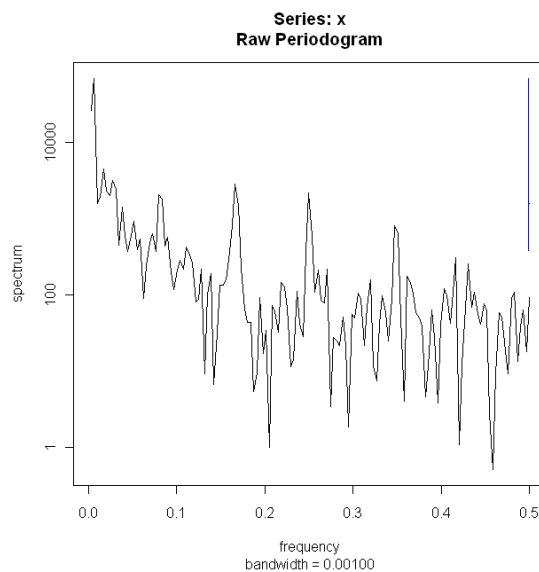
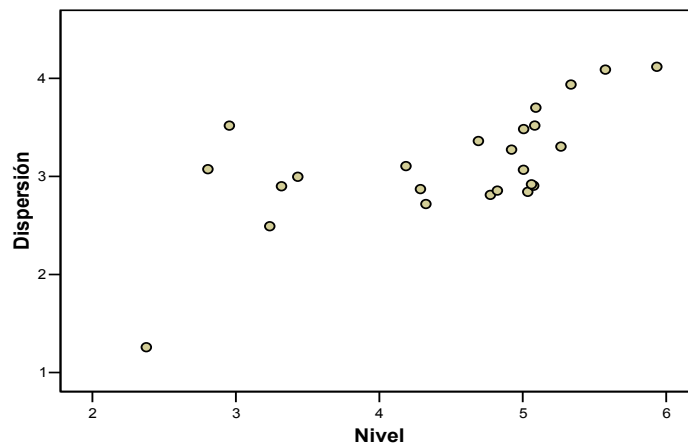


Grafico III.6

En el grafico III.6 podemos observar un pico aproximadamente en la frecuencia $0.02=1/50$, lo cual indica que es posible que nuestra serie presente una periodicidad (ciclo) de 50 meses.

Gráfico de dispersión por nivel de VAR00001 por YEAR_



* Gráfico de LN de dispersión por LN de nivel

Inclinación = .394 Potencia para transformación = .606

Grafico III.7

Según el grafico existe una relación lineal entre el nivel y la dispersión, lo cual indica que tenemos que transformar nuestra serie, la transformación a utilizar la escogeremos más adelante.

III.2 Identificación de memoria larga en la serie “Importación de Bienes de Consumo en el Perú (\$ mill)”

Ya en el grafico de correlograma hemos visto un indicio de memoria larga pues las autocorrelaciones decaen lentamente, sin embargo ahora utilizaremos la demás herramientas para corroborar esta característica en nuestra serie.

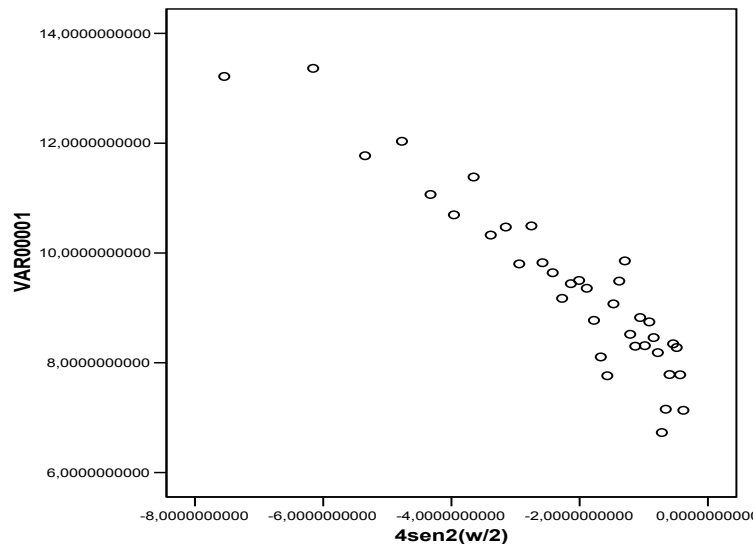


Grafico III.8

El gráfico III.8 (grafico del logaritmo de las frecuencias del periodograma vs $\ln(4\text{sen}^2(w/2))$), muestra un comportamiento lineal característica de series con memoria larga, ya tenemos dos argumentos para decir que nuestra serie presenta memoria larga. Veamos a continuación las docimas de hipótesis para contrastar la hipótesis de memoria larga.

La prueba de Geweke y Porter - Hudak (1983) para contrastar la hipótesis de memoria larga en la serie “Importación de Bienes de Consumo (\$ mill)”

Esta prueba contrasta las siguientes hipótesis:

H0: La serie presenta memoria corta ($d=0$)

H1: La serie presenta memoria larga ($d>0$)

El estadístico de prueba es

$$GPH = \frac{\hat{d} - d}{var(\hat{d})^{1/2}} \xrightarrow{dist} N(0, 1)$$

Para calcular el estadístico utilizamos el programa de software libre R, bajo la hipótesis nula ($d=0$), al calcular el estadístico del p-valor nos da 6.084586e-09 (con $d=0.841$ estimado por GPH) con lo cual se rechaza la hipótesis nula y decimos que nuestra serie presenta memoria larga.

La prueba propuesta por Elkin Castaño, Karoll Gómez y Santiago Gallón (2008) para contrastar la hipótesis de memoria larga en la serie “Importación de Bienes de Consumo (\$ mill)”

Esta prueba también contrasta las siguientes hipótesis:

H0: La serie presenta memoria corta ($d=0$)

H1: La serie presenta memoria larga ($d>0$)

El estadístico de prueba es $t_d = \frac{d}{se(d)}$, también calculamos el estadístico en el programa R el cual nos da como resultado un valor prácticamente cero (con $d=0.4991$ estimado por MMV) con lo cual rechazamos la hipótesis nula.

Ahora si podemos confirmar que la serie “Importación de Bienes de Consumo (\$ mill)” Presenta memoria larga.

III.3 Análisis de la serie “Importación de Bienes de Consumo (\$ mill)” bajo los modelos ARFIMA

Para realizar la aplicación a través de los modelos ARFIMA, se utilizó el software R.2.9.1, en este paquete se cargo la librería referida a la diferenciación fraccional “fracdiff”. Como hemos visto anteriormente en el análisis descriptivo nuestra serie necesita ser transformada y necesitamos estimar el parámetro de diferenciación fraccional y los parámetros autoregresivos y/o medias móviles; primero calculamos estimaciones para d , las estimaciones disponibles en R son: el método de regresión de la ecuación del periodograma (GPH), el método de regresión de la ecuación del periodograma suavizado (Sperio) y el Método de Máxima Verosimilitud (MMV). Los resultados obtenidos fueron los siguientes:

Metodo	"d" Estimado
MMV	0.4991628
GPH	0.8410965
Sperio	0.8354567

Tabla III.1

Se observa que el único parámetro estimado que cumple con las condiciones de estacionariedad es el calculado por MMV y, si bien es cierto en las docimas de hipótesis se vio que con el estadístico GPH se rechaza la hipótesis nula de memoria corta, no cumple con la condición de estacionariedad, por lo tanto para nuestro modelo utilizaremos el Método de Máxima Verosimilitud para el cálculo del parámetro de diferenciación fraccional d .

Lo que se hizo en este trabajo fue programar una rutina en el mismo software R (VER ANEXOS) para comparar varios modelos con distintas transformaciones y distintos parámetros de diferenciación fraccional, autoregresivos y medias móviles con su respectivo valor estadístico AKAIKE. Las transformaciones utilizadas fueron potencias a la -2, -0.8, ..., -0.4, -0.2, 0, 0.2, 0.4,...,2 (para la potencia cero se considero la transformación logarítmica), al observar los distintos modelos observamos que los que presentan un menor valor para el estadístico AKAIKE son los modelos con transformación logarítmica, más específicamente los modelos ARFIMA(1,d,3) y ARFIMA(3,d,2) con 43.00555 y 32.88657 valor AKAIKE respectivamente, por lo tanto la transformación usada será la logarítmica y escogemos el primer modelo para nuestra serie por el principio de parsimonia ARFIMA(1,d,3). Una ventaja del R es que

permite estimar todo los parámetros a la vez, tanto el parámetro de diferenciación fraccional como los parámetros autoregresivos y medias móviles.

En consecuencia nuestro modelo tendrá la siguiente forma:

MODELO ARFIMA (1, d, 3)

$$(1 - \phi_1 B) \nabla^d y'_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3) a_t$$

Finalmente el modelo estimado es el siguiente:

$$(1 + 0.09782B) \nabla^{0.47} y'_t = (1 + 0.39905B + 0.25735B^2 + 0.32783B^3) a_t$$

Si tenemos en cuenta la sumatoria del operador de diferenciación fraccional:

$$\nabla^{0.47} = \sum_{k=0}^{\infty} \binom{0.47}{k} (-B)^k = 1 - 0.47B - \frac{0.47(1-0.47)B^2}{2!} - \frac{0.47(1-0.47)(2-0.47)B^3}{3!} - \dots$$

Y cuando desagreguemos la ecuación de nuestro modelo teniendo en cuenta esta sumatoria, se podrá ver que el modelo contiene a todos los valores de la serie demostrando así ser más consistente que otros modelos como los modelos ARIMA.

Validación y Pronósticos

Para la validación del modelo debemos utilizar los residuales, sin embargo en el paquete “fracdiff” del programa R en la actualidad aun no están implementados los comandos referidos a residuales y predicciones; ante esta situación se hizo una rutina de programación en R para poder obtener los residuales y pronósticos (VER ANEXO), al analizar los residuales obtenemos lo siguiente:

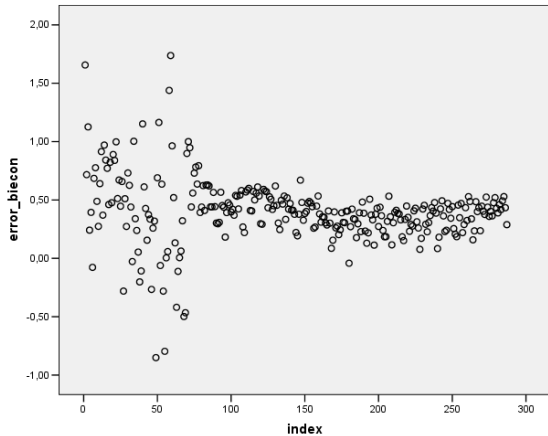


Grafico III.9

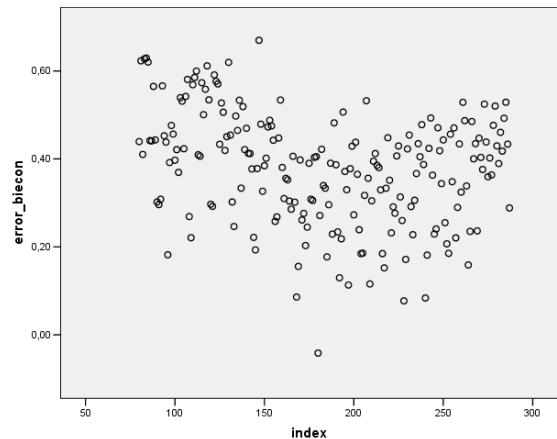


Grafico III.10

Al observar los residuales con respecto al tiempo podemos ver que se muestran aleatorios, aunque en el grafico III.9 los residuales parecen mostrar un comportamiento esto no es así, sucede que el rango de los residuales es más amplio desde el primer residual hasta aproximadamente el residual 79, en el grafico III.10 vemos que a partir del residual 80 el rango disminuye y los errores se estabilizan pero siguen manteniéndose aleatorios.

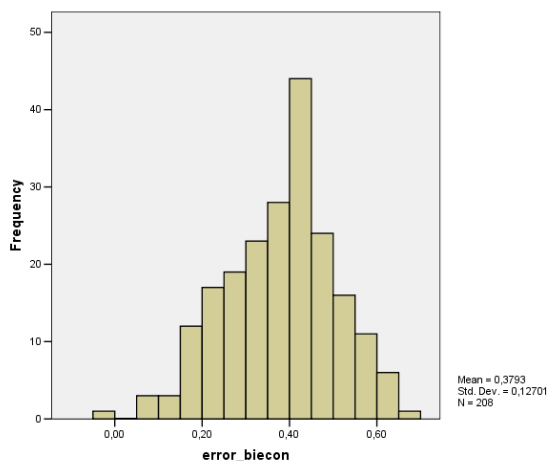


Grafico III.11

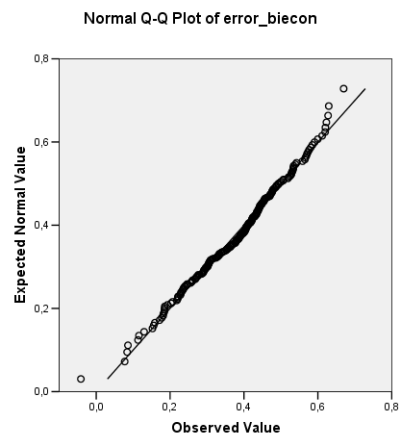


Grafico III.12

Podemos observar que los residuos se aproximan a una distribución normal, el grafico III.11 es casi simétrico y el grafico III.12 muestra en los puntos un comportamiento muy aproximado a la línea.

Para los pronósticos se hicieron dos rutinas (VER ANEXOS), una considerando los valores estimados o ajustados y otra rutina considerando los valores reales, al final los pronósticos calculados considerando los valores reales se ajustan mejor a los reales:

PRONOSTICOS		
Meses 2009	Datos reales	ARFIMA(1,0.47,3)
Enero	314.686246	317.5755326
Febrero	303.630034	292.861717
Marzo	311.116858	274.7947861
Abril	305.777572	261.1806425
Mayo	291.545635	250.237921
Junio	307.075056	241.1340058
Julio	319.213035	233.3684974

Tabla III.2

Se observa que los pronósticos del modelo ARFIMA para los meses de enero, febrero y marzo son los más próximos, a partir del mes de abril se van alejando del valor real (aunque el valor para los meses de abril y mayo también son buenos). Debemos señalar que hemos considerado para los pronósticos los términos de los últimos tres errores estimados (286, 287 y 288) ya que nuestro modelo tiene tres términos medias móviles, inicialmente calculamos los pronósticos (a partir del 290), sin considerar los términos medias móviles en el modelo, sin embargo, pudimos evidenciar que al considerar los términos medias móviles los pronósticos mejoran mucho.

III.4 Comparación con el modelo ARIMA

Al estimar el mismo modelo por el método ARIMA (1, d=1,3), tenemos lo siguiente:

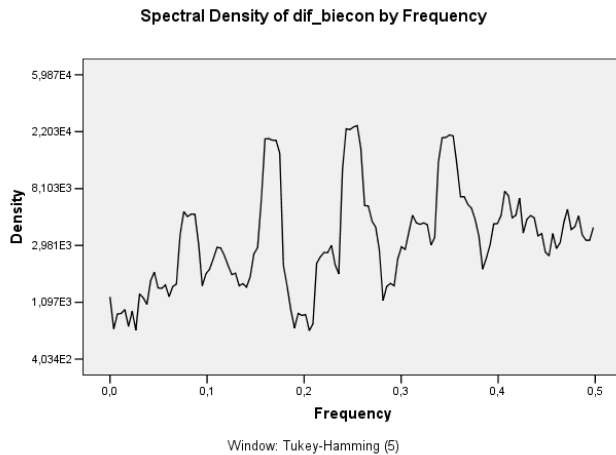


Grafico III.13

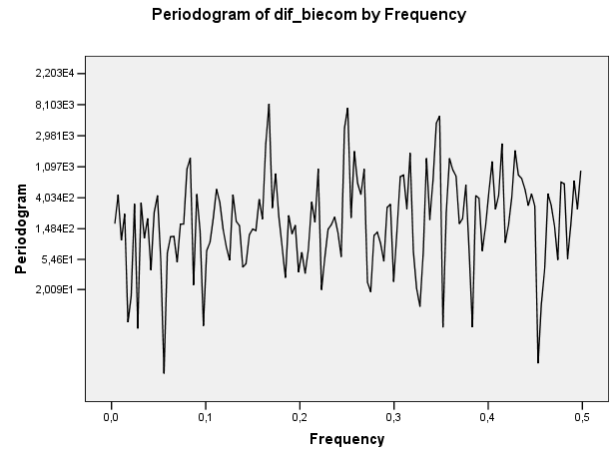


Grafico III.14

En los gráficos se aprecia una de las desventajas de los modelos ARIMA cuando se trabaja con series de memoria larga, vemos que la función de densidad espectral y el periodograma después de haber diferenciado la serie ($d=1$), tiende a anularse en el origen y está dominada por las frecuencias altas.

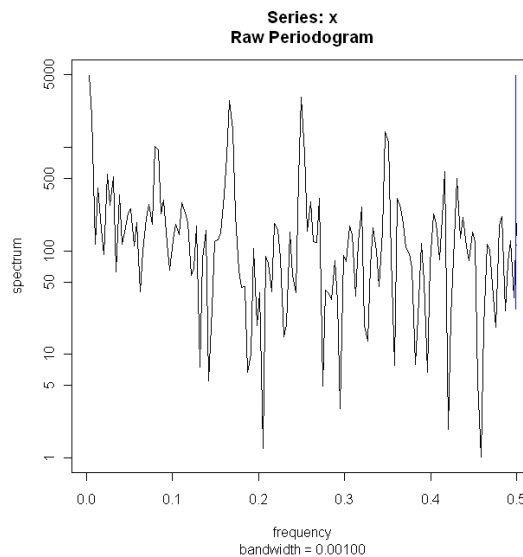


Grafico III.15

Mientras que en el grafico III.15 se observa que el periodograma de nuestro modelo ARFIMA estimado mantiene las componentes a largo plazo al tener valores más altos para las frecuencias cercanas a cero.

Finalmente en la tabla podemos comparar los pronósticos obtenidos por los métodos ARIMA y ARFIMA con los valores reales.

PRONOSTICOS			
Meses 2009	Datos reales	ARFIMA(1,0.47,3)	ARIMA(1,1,3)
Enero	314.686246	317.5755326	390.83725
Febrero	303.630034	292.861717	397.88356
Marzo	311.116858	274.7947861	399.89727
Abril	305.777572	261.1806425	397.98469
Mayo	291.545635	250.237921	399.80078
Junio	307.075056	241.1340058	398.07591
Julio	319.213035	233.3684974	399.71377

Tabla III.3

Observamos como los pronósticos del modelo ARIMA tienden a incrementarse y mantenerse casi constantes, por el contrario los pronósticos del modelo ARFIMA mantienen un comportamiento próximo a los valores reales, observándose mejor en el grafico III.16 de los valores reales y los valores ajustados por el modelo ARFIMA (1,0.47, 3).

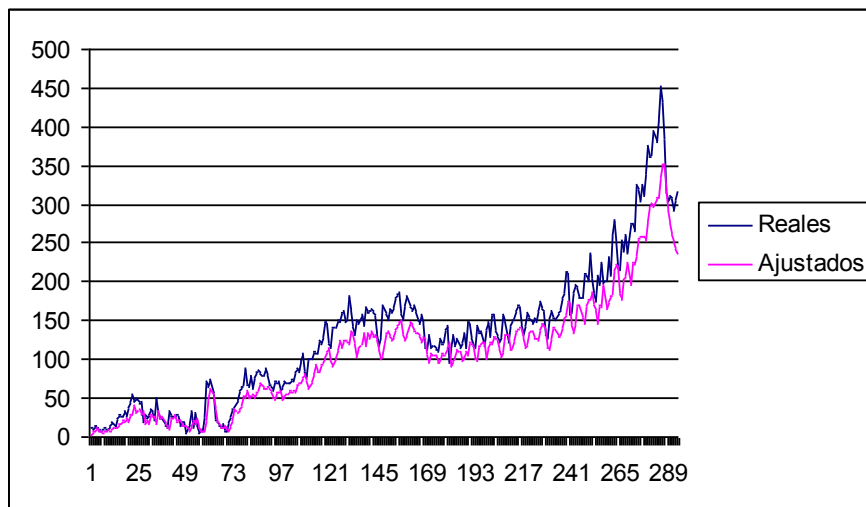


Grafico III.16

Se debe recalcar que los modelos ARFIMA al no estar asociados a ninguna representación de dimensión finita presentan un problema de truncamiento asociado a la representación autoregresiva para predicción.

CONCLUSIONES

Las docimas de hipótesis presentadas en este trabajo, constataron la presencia de memoria larga en la serie de tiempo presentada y constituyen un conjunto de herramientas relevantes a las ya conocidas para detectar memoria larga en series de tiempo.

El cálculo de las estimaciones es más preciso con los modelos ARFIMA a comparación de los modelos ARIMA, ya que para calcular cada valor pronosticado se utilizan todo los valores pasados hasta el momento t .

Se debe considerar un corto horizonte de pronóstico, sobre todo cuando el modelo está compuesto por términos medias móviles, teniendo en cuenta que no se cuenta con estos valores a partir del tiempo $t+1$, los pronósticos serán casi exactos solo hasta algunos pasos hacia adelante.

Se debe contar con una cantidad numerosa de datos para poder trabajar con los modelos ARFIMA en relación a la característica de memoria larga, además los pronósticos serán más exactos.

La crisis económica mundial a finales del 2008 afectó a la disminución de importaciones de bienes de consumo en el Perú, el modelo ARFIMA estimado logro captar este comportamiento a diferencia del modelo ARIMA cuyo comportamiento es creciente.

BIBLIOGRAFÍA

1. Brockwell P.–Davis R.A. “Time series: Theory and Methods” (1987)
2. Banco Central de Reservas del Perú. Notas de Estudios, (Setiembre, Octubre 2009)
3. Castaño Elkin, Gallon Santiago y Gómez Karoll “Una nueva prueba para el parámetro de diferenciación fraccional” Revista Colombiana de Estadística, volumen 31, no 1, pp. 67 a 84. (Junio 2008)
4. Domínguez C. Melina “Modelos ARFIMA una alternativa para modelar series de memoria larga” Tesis de Licenciatura. Facultad de Ciencias Matemáticas (2000)
5. Granger, C.W.J y R. Joyeux “An Introduction to Long Memory Time Series Models and Fractional Differencing”. Journal of Time Series Analysis, 1, 15-29. (1980)
6. Hosking, J.R.M. “Fractional Differencing”. Biometrika, 68, 165-176 (1981)
7. Paradis Emmanuel “R para principiantes” Institut des Sciences de l’E´volution Universit Montpellier II (traducido por Ahumada Jorge A. (2003)
8. Peña Daniel. “Análisis de Series Temporales” Alianza Editorial (2005)
9. Peiró Amado - Uriel Ezequiel “Introducción al análisis de Series Temporales”. Editorial A.C – Madrid (2000)

10. Pérez Espartero Ana “Modelos con memoria larga para la media: Procesos ARFIMA” Dpto. Economía Aplicada (Estadística y Econometría) Universidad de Valladolid
11. Reisen Valdério A. “Arfima –O Modelo Arima Para O d Fraccionario” 6ª Escola de Seires Temporais e Econometria(1995)
12. Reisen V. and B. Abraham “Prediction of long Memory Time Series Models: A simulation study and a application” Univeristy of Waterloo (1998)
13. Seier Edith “¿Qué es el análisis espectral y Como se Aplica? “ UNMSM-Notas didácticas (1992)
14. Spremolla Alexandra “Persistencia en el desempleo de Uruguay” Instituto de Economía, Pontificia Universidad Católica de Chile (2007)
15. Shumway Robert H. Stoffer David S. “Time Series Analysis and Its Applications” (2000)
16. SUNAT Nota de Prensa N° 186-2009

ANEXOS

Rutinas en el programa R utilizadas para la aplicación

#Carga la serie calcula gráficos de la serie autocorrelaciones y espectro

```
serie1=read.table("serie bienes consumo BCRP.txt",header=T)
serie1
serie2=serie1[1:288,]
serie2
plot.ts(serie2[,2])
#función de autocorrelación
acf(serie2[,2],lag.max=70)
#espectro en el dominio de la frecuencia
spectrum(serie2[,2])
library(fracdiff)
```

#calcula el d por MMV y calcula el p-valor para la docima de hipotesis de memoria larga

```
mFD = fracdiff( log(serie2[,2]))
coef(mFD)
confint(mFD)
mFD$d
mFD$ar
mFD$ma
estadistico1=mFD$d/mFD$stderror.d
pvalor=1-pnorm(estadistico1)
pvalor
```

#calcula d con estimador GPH y calcula el p-valor para la docima de hipotesis de memoria larga

```
mGPH = fdGPH(log(serie2[,2]))
mGPH$d
mGPH$sd.as
mGPH$sd.reg
estadistico2=mGPH$d/mGPH$sd.reg
pvalor=1-pnorm(estadistico2)
```

pvalor

#diferencia y correlaciones por MMV

```
r <- diffseries(serie2[,2], d = mFD$d)
```

```
r
```

```
acf(r)
```

```
pacf(r)
```

#diferencia y correlaciones por GPH

```
g <- diffseries(serie2[,2], d = mGPH$d)
```

```
g
```

```
acf(g)
```

#diferencia y correlaciones por SPERIO

```
spm = fdSperio(serie2[,2])
```

```
str(spm, digits=6)
```

```
s <- diffseries(serie2[,2], d = spm$d)
```

```
s
```

```
acf(s)
```

```
pacf(s)
```

```
confint(spm$d)
```

#Modelo Completo Máxima Verosimilitud - Rutina Que Calcula Varios Modelos Y Sus Respectivos Valores Del Estadístico Akaike

```
iter=0
```

```
potencia=seq(-2,2,0.2)
```

```
longitud=length(potencia)*16
```

```
longitud
```

```
akaike=rep(0,longitud)
```

```
AR=rep(0,longitud)
```

```
MA=rep(0,longitud)
```

```
pot=rep(0,longitud)
```

```
for (z in potencia)
```

```
for (i in 0:3)
```

```
for (j in 0:3)
```

```

{
iter=iter+1
if (abs(z) > 0)
modelo=fracdiff( serie2[,2]^z, nar = i, nma = j)
if (z == 0)
modelo=fracdiff( log(serie2[,2]), nar = i, nma = j)
pot[iter]=z
AR[iter]=i
MA[iter]=j
akaike[iter]=AIC(modelo)
}
consolidado=data.frame(pot,AR,MA,akaike)
consolidado

```

#MODELO1 log ar=1 ma=3 Estimado

```

modelo1=fracdiff( log(serie2[,2]), nar = 1, nma = 3)
coef(modelo1)
confint(modelo1)
diferencia= diffseries(log(serie2[,2]), d = modelo1$d)
diferencia
par(mfrow=c(2,1))
acf(diferencia,lag.max=100)
pacf(diferencia,lag.max=100)
par(mfrow=c(1,1))
summary(modelo1)
vcov(modelo1)
AIC(modelo1)

```

#Predicción y residuales en base a las observaciones anteriores

```

n=length(serie2[,1])
YLOG=log(serie2[,2])
YLOGP=rep(0,n+12)
ERR=rep(0,n+12)
ALFA=rep(0,n+12)
for (k in 1:length(ALFA))
{

```

```

ALFA[k]=((-1)^(k+1))*(choose(modelo1$d,k)+modelo1$ar*choose(modelo1$d,k-1))
}
YLOGP[2]=ALFA[1]*YLOG[1]
ERR[2]=YLOG[2]-YLOGP[2]
YLOGP[3]=ALFA[1]*YLOG[2]+ALFA[2]*YLOG[1]-modelo1$ma[1]*ERR[2]
ERR[3]=YLOG[3]-YLOGP[3]
YLOGP[4]=ALFA[1]*YLOG[3]+ALFA[2]*YLOG[2]+ALFA[3]*YLOG[1]-
modelo1$ma[1]*ERR[3]-modelo1$ma[2]*ERR[2]
ERR[4]=YLOG[4]-YLOGP[4]
YLOGP[5]=ALFA[1]*YLOG[4]+ALFA[2]*YLOG[3]+ALFA[3]*YLOG[2]+ALFA[4]*YLOG[1]
-modelo1$ma[1]*ERR[4]-modelo1$ma[2]*ERR[3]-modelo1$ma[3]*ERR[2]
ERR[5]=YLOG[5]-YLOGP[5]
for (i in 5:(n+1))
{
suma=0
for (j in 1:(i-1))
{
suma=suma+ALFA[j]*YLOG[i-j]
}
YLOGP[i]=suma-modelo1$ma[1]*ERR[i-1]-modelo1$ma[2]*ERR[i-2]-
modelo1$ma[3]*ERR[i-3]
ERR[i]=YLOG[i]-YLOGP[i]
}
for (i in (n+2):(n+12))
{
suma=-modelo1$ma[1]*ERR[n]-modelo1$ma[2]*ERR[n-1]-modelo1$ma[3]*ERR[n-2]
for (j in 1:n)
{
suma=suma+ALFA[i-j]*YLOG[j]
}

for (j in (n+1):(i-1))
{
suma=suma+ALFA[i-j]*YLOGP[j]
}
YLOGP[i]=suma

```

```
}  
consolidadoobs=data.frame(YLOGP,ERR)  
spectrum(YLOGP)  
  
consolidadofinal=data.frame(consolidadopred,consolidadoobs)
```

Importación de bienes de consumo en el Perú (mill \$)												
Años	Meses											
	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Setiembre	Octubre	Noviembre	Diciembre
1985	12.12	13.13	8.5	14.99	11.34	10.18	6.48	9.34	11.59	9.48	9.41	13.21
1986	18.09	12.49	24.27	27.74	27.49	25.34	34.66	26.99	37.79	44.35	55.6	44.56
1987	50.77	42.83	45.57	18.97	28.95	23.52	29.45	35.29	31.18	21.69	49.45	30.67
1988	25.08	23.49	13.82	11.71	33.21	26.36	26.57	27.17	28.04	25.81	13.26	18.3
1989	18.68	5.31	17.77	34.37	12.34	30.58	15.29	5.76	9.28	7.39	26.18	71.42
1990	65.77	73.98	54.13	22.37	18.64	15.18	12.65	16.25	7.76	6.25	19.76	25.53
1991	36.053	39.241	48.511	60.862	65.776	67.007	87.40654	67.083	65.735	78.746	62.073	76.19623
1992	84.33051	86.4112	80.46064	79.525	88.2896	78.3617	67.75016	64.8667	60.4124	72.7868	69.5736	71.1782
1993	59.132583	65.577801	70.817262	68.99547	70.366153	74.116014	71.310896	83.142873	88.176929	83.410215	98.35434	107.72099
1994	82.792864	76.049613	99.509349	101.40063	110.92675	106.86925	107.02553	125.3672	118.83508	129.94345	148.97204	146.0942
1995	120.80623	113.74882	140.81076	140.78919	149.77627	147.18131	161.35547	162.59565	148.63895	154.26824	181.95296	162.97405
1996	142.14471	130.93951	150.39357	145.43447	158.80208	144.4101	168.4652	161.20005	163.86728	163.89998	160.86851	156.60728
1997	130.78413	118.05564	129.93188	169.27518	158.78504	149.81884	164.65474	160.47342	167.81041	178.25807	184.32318	187.34747
1998	158.98431	150.89395	169.38281	182.52973	168.08763	163.31737	168.82858	161.32905	150.20872	145.06763	158.73137	144.89978
1999	114.73515	113.58825	131.73252	114.33977	116.86497	113.22453	111.15945	126.46565	119.97746	122.26965	139.57116	143.70908
2000	95.869716	115.99193	131.74475	117.81658	127.94776	114.96087	121.58799	132.8761	114.26251	147.25911	146.7698	127.12464
2001	116.50615	115.58462	143.99565	134.57514	136.97067	118.27853	138.86302	148.40083	128.7536	158.63092	156.85981	137.48307
2002	129.99302	121.15137	128.0182	158.54473	147.55501	123.1969	143.81747	151.06901	152.65594	161.89078	169.817	166.422
2003	143.48862	130.80531	144.52575	159.22208	153.68623	145.46219	153.62504	147.3633	162.28971	175.32668	164.66798	160.82413
2004	131.33032	127.9793	153.78486	162.38583	152.37993	151.4577	160.26124	165.06301	179.53847	187.38853	213.06423	210.49832
2005	158.27106	159.27789	186.09631	196.42263	191.0206	180.19192	178.26708	210.97893	207.19612	201.9581	235.7764	202.3033
2006	184.24192	173.73879	207.11597	195.14539	225.3013	197.79288	200.5854	233.05292	208.25191	260.32285	278.81775	251.7356
2007	218.13641	214.98751	254.67106	240.13626	261.9572	235.90929	275.88484	276.31575	266.10925	325.9	318.3083	303.44053
2008	324.669	310.29717	340.90606	374.59281	361.76697	363.60971	393.84967	380.91045	412.1033	451.04421	430.23465	383.16847
2009	314.68625	303.63003	311.11686	305.77757	291.54564	307.07506	319.21304					

Fuente: Banco Central de Reserva del Perú - BCRP